GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 09:04:18 ; Search time 2548.21 Seconds
(without alignments)
4179.050 Million cell updates/sec

Title: US-09-863-823-5
Perfect score: 789
Sequence: 1 atggcatggaagagcagtgt.....acagtgaaacagctctatga 789

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	671.4	85.1	822	10	BI764721	BI764721 603051058
2	504.6	64.0	1367	11	AK009118	AK009118 Mus muscu
3	487.2	61.7	1317	11	AK008028	AK008028 Mus muscu
4	435.4	55.2	797	10	BG970132	BG970132 602839263
5	425.6	53.9	646	9	AA986919	AA986919 uc80b02.y
6	415.8	52.7	964	10	BF785067	BF785067 602108226
7	402.6	51.0	538	10	BM445724	BM445724 IIL15A10.
8	382.2	48.4	936	10	BF788117	BF788117 602113341
9	381.8	48.4	543	9	AA500290	AA500290 vi97g08.r
10	336.2	42.6	657	9	AI747948	AI747948 ul04c10.y
11	306.6	38.9	445	9	AI956638	AI956638 ul79f12.y
12	295.6	37.5	487	9	AI615364	AI615364 vi97g08.y
13	241.2	30.6	514	9	AW211197	AW211197 ul25h03.y
14	215.4	27.3	437	9	BB845854	BB845854 BB845854
c 15	189.8	24.1	515	12	AZ864346	AZ864346 2M0173F20
c 16	160	20.3	504	9	AW235098	AW235098 xn18e10.x
c 17	156.8	19.9	175	9	AW844313	AW844313 RC2-CN005
18	131.8	16.7	535	9	AW107893	AW107893 um18d09.x
19	108	13.7	504	10	W82090	W82090 me98e12.r1
20	106.4	13.5	320	9	AI425843	AI425843 me98e12.y
21	90.6	11.5	125	10	BM445552	BM445552 IIL12H10.
22	61.2	7.8	545	10	BM026157	BM026157 fu75c12.y
c 23	52	6.6	572	9	AI790737	AI790737 ul04c10.x
c 24	51.4	6.5	590	10	BM025920	BM025920 fu75c12.x
25	43.8	5.6	151	9	AA489845	AA489845 ab04g06.s
c 26	43.4	5.5	522	9	AW107893	AW107893 um18d09.x
27	40.6	5.1	432	9	AA772324	AA772324 ai43a08.s
28	40.6	5.1	780	12	CNS04I4Z	AL291788 Tetraodon
c 29	40.2	5.1	393	10	W94303	W94303 zd80f12.s1
30	40.2	5.1	411	12	AZ979833	AZ979833 2M0256107
31	39.8	5.0	435	9	AI672771	AI672771 we58f05.x
32	39.2	5.0	433	9	AI032799	AI032799 ow17c10.x
33	39.2	5.0	513	9	AI160022	AI160022 qc91b04.x
c 34	39.2	5.0	523	10	BF063348	BF063348 7h88e07.x
35	39.2	5.0	637	10	BE350315	BE350315 ht13b10.x
36	39.2	5.0	670	10	BM171749	BM171749 imageqc_5
37	39.2	5.0	816	9	AU117395	AU117395 AU117395
38	39	4.9	252	10	BF785567	BF785567 602112349
39	39	4.9	474	9	BB818563	BB818563 BB818563
c 40	39	4.9	610	12	FR0021997	AL014868 F.rubripe
41	39	4.9	642	9	BB652201	BB652201 BB652201
42	39	4.9	665	9	BB614047	BB614047 BB614047
43	39	4.9	676	9	BB465902	BB465902 BB465902
44	39	4.9	940	11	AK020591	AK020591 Mus muscu
45	38.2	4.8	478	9	AA844204	AA844204 ai91h02.s

ALIGNMENTS

RESULT 1
BI764721
LOCUS BI764721 822 bp mRNA linear EST 25 SEP 2001
DEFINITION 603051058F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5190878 5', mRNA sequence.
ACCESSION BI764721
VERSION BI764721.1 GI:15756299
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11477 row: d column: 15
High quality sequence stop: 821.
FEATURES Location/Qualifiers
source 1..822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5190878"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon, 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.4 kb, insert size range 1 3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIB_MGC Library."
BASE COUNT 251 a 166 c 195 g 210 t
ORIGIN

Query Match 85.1%; Score 671.4; DB 10; Length 822;
Best Local Similarity 97.4%; Pred. No. 1.2e-180;
Matches 754; Conservative 3; Mismatches 9; Indels 8; Gaps 7;

```
Qy       9 qaagaqcaqtqtcataatgcaratqqqaaqatttcttctcttaqtaattttatttctqcc 68
           ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db       1 GAAGAGCAGTGTCATAATGCAAATGGGAAGATTTCTTCTCTTAGTAATTTTATTTCTGCC 60

Qy      69 acqtqaqatqacaaqttctqttttaactqtqaatqqtaaaactqaqaaactatatcctqqa 128
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      61 ACGTGAGATGACAAGTTC-GTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTGGA 119

Qy     129 tactacacctqqctcccaaqcatctctqatatqtqctgttcaaaaccacaccaqaqaqqa 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 TACTACACCTGGCTCCCAAGCATCTCTGATATGTGCTGTTCAAAACCACACCAGAGAGGA 179

Qy     189 aqaactqctctqqtaccqaqaqqaqqqqaqagtqqatttqaaatctgsaaacaaatcaa 248
           |||||||||||||||||||||||||||||||| |||||||||||:||||||:|||||
Db     180 AGAACTGCTCTGGTACCGAGAGGAGGGGAGAGTCCA-TTCAAATCTGGAAACAAAATCAA 238

Qy     249 ttccaqctctqtctqtqtctcttccatcaqtqaaaatqacaaccqqaatcaqctttacctq 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     239 TTCCAGCTCTGTCTGTGTCTCTTCCATCAGTGAAAATGACAACGGAATCAGCTTTACCTG 298

Qy     309 caqqctqqqqaqqqatcaqtcaqtccqtqtccqttcqqtqqtqctqaatgttacttttcctcc 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 CAGGCTGGGGAGGGATCAGTCCGTGTCCGTTCGGTGGTGCTGAATGTTACTTTTCCTCC 358

Qy     369 tctcctaaqtqqaaacqacttccaaacaqttqaqqaaqqcaqtaatgtgaaqttggtttq 428
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     359 TCTCCTAAGTGGAAACGACTTCCAAACAGTTGAGGAAGGCAGTAATGTGAAG-TGGTTTG 417

Qy     429 caatqtqaaaqccaaccccccaqqctcaaatqatqtqqtacaaaaacagtaqtctcctcqa 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     418 CAATGTGAAAGCCAACCCCCAGGCTCAAATGATGTGGTACAAAAACAGTAGTCTCCTCGA 477

Qy     489 tttaqaqaaaaqccqtcaccaaatccaaccaqacaaqtqaqtctttcaqctqtcaatcac 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     478 TTTAGAGAAAAGCCGTCACCAAATCCAACAGACAAGTGAGTCTTTTCAGCTGTCAATCAC 537

Qy     549 caaaqtcqaqaaqcctqacaaccqqaacctacaqttgtatttqcaaaqtcatctctqaaaac 608
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     538 CAAAGTCGAGAAGCCTGACAACCGGAACCTACAGTTGTATTGCAAAGTCATCTCTGAAAAC 597

Qy     609 qqaqaqcttqqactttcaccqtqattqttaaaqataaaactqtqqqtqtaccaatagagcc 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     598 GGACAGCTTGGACTTTCACCTGA-TGTTAAAGATAAAACTGTGGGTGTACCAATAGAGCC 656

Qy     669 cattattgtqcatqtqttqtqatctttctqacattqtqctttqqactqattqctaqaaq 728
           |||||||| |||||||||||||||||||||||||||||| |      |||||||||
Db     657 CATTATTG-TGCATGTGTTGTGATCTTTCTGACATTGTGCTT--GGACTGATGCTAGAAG 713

Qy     729 aaaqaaaataatqaaqctctqcatqaaqgataaaqaccctcacaqtgaaacaq 781
           |||||| |||||||||||||| |||||||||||| |||||||||||||||||
Db     714 AAAGAACATAATGAAGCTCTGCTTGAAGGATAAAGA-CCTCACAGTGAAACAG 765
```

RESULT 2
AK009118
LOCUS AK009118 1367 bp mRNA linear HTC 19 JAN 2002
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310009K08:Immunoglobulin domain containing protein, full insert sequence.
ACCESSION AK009118
VERSION AK009118.1 GI:12843711
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2310003K08.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap trapper selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,F., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1367)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schrimi,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45 503 9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCICIIIIIIIIIIIIIIIVN 3'], cDNA was prepared by using trehalose thermo activated reverse transcriptase and subsequently enriched for full length by cap trapper. cDNA went

through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3'], cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end. XhoI, 3' end. SstI.
Host: SOLR.

```
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1898129"
                     /db_xref="taxon:10090"
                     /clone="2310003K08"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             24. .809
                     /note="Immunoglobulin domain containing protein
                     data source:Pfam, source key:PF00047, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB26083.1"
                     /db_xref="GI:12843712"
                     /translation="MVWKITGPLQACQLLLVVLSLPQGRTSSVLTVNGRTENYILDTQ
                     HGVQASLECAVQNHTEDEELLWYREDGIVDLKNGNKINISSVCVSPINESDNGVRFTC
                     KLQRDQTVSVTVVLNVTFPPLLSGNGFQTVEENSDVSLVCNVKSNPQAQMMWYKNNSA
                     LVLEKGRHQIHQTRESFQLSITKVKKSDNGTYSCIASSSLKMETMDFHLLVKDKVFVM
                     PAEPIIAACVVVVLTMAFALFSPRKRIMKLCGKKNDPNSETAL"
BASE COUNT       417 a    267 c    298 g    385 t
ORIGIN
```

Query Match 64.0%; Score 504.6; DB 11; Length 1367;
Best Local Similarity 79.5%; Pred. No. 5.6e-133;
Matches 594; Conservative 2; Mismatches 151; Indels 0; Gaps 0;

```
Qy      43 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 102
           ||||||||||||  ||||||  | ||||||  | ||  |||||  ||||||||||||||||||||
Db      63 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 122

Qy     103 ggtaaaactgagaactatatcctggatactacacctggctcccaagcatctctgatatgt 162
           ||||  ||||||||||||||||  |||||  |||   ||  |||   ||||||||||||||  ||||
Db     123 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 182

Qy     163 gctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggggagagtg 222
           ||||||||||||||||||   ||  ||||||||  |||||||||  ||||  ||  ||  |  |||
Db     183 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 242

Qy     223 gatttgaaatctgaaaacaamatcaattccagctctgtctgtgtctcttccatcagtgaa 282
           |||  |||||   ||:||||||:|||||    |||  |||||  |||||||||  ||||||   |||
Db     243 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 302

Qy     283 aatgacaacggaatcagctttacctgcaggctggggagggatcagtccgtgtccgtttcg 342
           |  ||||||  |||  ||  ||||||||||||  |||    |  |  ||||||  |  ||||||||  |
Db     303 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 362

Qy     343 gtggtgctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgag 402
           ||  ||||||||  ||||  ||||||||||  ||||||||  ||  |  |||||||||  ||  |||
Db     363 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 422

Qy     403 gaaggcagtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatg 462
           |||  |||  ||||  |  ||||||||||  ||||||  ||||||  ||  ||  |||||||||
Db     423 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 482

Qy     463 tggtacaaaaacagtagtctcctgatttagagaaaagccgtcaccaaatccaacagaca 522
           |||||  |||||||  ||  |  |  |  ||||||||||  ||||||||||||||||  ||||||
Db     483 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACA 542

Qy     523 agtgagtcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagt 582
           ||  ||||||||||  ||||  ||||||||||||  ||||  |||||||  ||  ||||||||
Db     543 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 602
```

```
Qy     583 tgtattgcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagat 642
           ||||||||  |  ||||||||||||||  |||||  |  |||||||  ||||||  ||||||||||
Db     603 TGTATTGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGAC 662

Qy     643 aaaactgtgggtgtaccaatagagcccattattgctgcatgtgttgtgatctttctgaca 702
           |||  |  |  |   |  ||   |||||||||  ||||||||  |||||  |||  ||  |  |||||
Db     663 AAACTTTTTGTCATGCCCCAGAGCCCATCATTGCTGCTTGTGTCCTGGCCTGACG 722

Qy     703 ttgtgctttggactgattgctagaagaaagaaaaataatgaagctctgcatgaaggataaa 762
           ||   ||||  |||  ||  ||||  ||  ||  |  ||||||||||||||||   |||  |  ||
Db     723 ATGGCTTTGCCCTGTTTTCTAGGAGGAAAAGAATAATGAAGCTCTGTGGGAAAAAGAAT 782

Qy     763 gacgctcacagtgaaacagctctatga 789
           |||||  |   ||||||||||||||  |
Db     783 GACCCGAATTCAGAAACAGCTCTATAA 809
```

```
RESULT   3
AK008028
LOCUS       AK008028              1317 bp    mRNA    linear   HTC 14-JAN-2002
DEFINITION  Mus musculus adult male small intestine cDNA, RIKEN full-length
            enriched library, clone:2010002A20:Immunoglobulin domain containing
            protein, full insert sequence.
ACCESSION   AK008028
VERSION     AK008028.1  GI:12841960
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to
            mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2010002A20.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1317)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
```

```
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,T.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 20.0. Second strand cDNA was prepared with the primer adapter
            of sequence[5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end. XhoI, 3' end. SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1317
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1894220"
                     /db_xref="taxon:10090"
                     /clone="2010002A20"
                     /sex="male"
                     /tissue_type="small intestine"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             85..846
                     /note="Immunoglobulin domain containing protein
                     data source:Pfam, source key PF00047, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB25419.1"
                     /db_xref="GI:12841961"
                     /translation="MVWKITGPLQACQLLLVVLSLPQGRTSSVLIVNGRTENYILDTQ
                     HGVQASLECAVQNHTEDEELLWYREDGIVDLKNGNKINISSVCVSPINESDNGVRFTC
                     KLQRDQTVSVTVVLNVTFPPLLSGNGFQTVEENSDVSLVCNVKSNPQAQMMWYKNNSA
                     LVLFKGRHQIPQTPFSEQLSLITVKKSDNGIYSCIASSSLAMPIMDFHLAVKDNVFVM
                     PAEPIIAACVVVVLTMAFALFSRKKRIMKFLWEKE"
     polyA_signal    1300..1305
                     /note="putative"
     polyA_site      1317
                     /note="putative"
BASE COUNT      401 a    252 c    304 g    460 t
ORIGIN


  Query Match             61.7%;  Score 487.2;  DB 11;  Length 1317;
  Best Local Similarity   78.9%;  Pred. No. 5.1e-128;
  Matches  590;  Conservative    2,  Mismatches  155,  Indels    1,  Gaps    1;

Qy      43 cttctcttagtaattttatttctgccaagtgagatgacaagttctgttttaaactgtgaat 102
           ||||||||||||  ||||||  | ||||||  | ||  |||||  ||||||||||||||||||
Db     124 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 183

Qy     103 qqtaaaactgaqaactatatcctggatactacacctqqctccaaqcatctctgatatgt 162
           |||| ||||||||||||||| ||||| |||  || |||   ||||||||||||  ||||
Db     184 GGTAGAACTGAGAACTATATCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 243

Qy     163 gctgttcaaaaccacaccgaggatgaagaacttcttctgctacagagaagatggaatagta 222
           ||||||||||||||||||   || ||||||| |||||||| |||| || || | |||
Db     244 CCTCTTCAAAACCACACCGAGGATGAAGAACTTCTTCTGCTACAGAGAAGATGGAATAGTA 303

Qy     223 gatttgaaatctgaaaacaaMatcaattcagctctgtctgtgtctcttccatcaqtqaa 282
           ||| ||||| ||:|||||:|||||   ||| ||||| |||||||| |||||| |||
Db     304 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 363

Qy     283 aatgacaacqgaatcaqctttacctqcaqqctgqqqaqqgatcaqtccqtqtccqtttcq 342
           | |||||| ||| || |||||||||||| |||   | | |||||| | |||||||||| |
Db     364 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 423

Qy     343 gtqqtqctqaatqttacttttcctcctctcttaaqtqqcaatqqcttccaaaccqttqaq 402
           || |||||||| ||||| |||||||||||| |||||||| || | |||||||||| || |||
Db     424 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 483

Qy     403 qaaggcaqtaatgtqaaqttqqtttgcaatqtqaaaaccaaccccccaqqctcaaatqatq 462
           |||  |||  |||| |  |||||||||||| |||||| |||||| || || |||||||||
Db     484 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 543

Qy     463 tqqtacaaaaacaqtaqtctcctqqatttaqaqaaaaqccqtcaccaaatccaaccaqaca 522
           ||||| |||||||  ||  | | | |||||||||| ||||||||||||||  ||||||
Db     544 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCCCCAGACA 603

Qy     523 aqtqaqtctttttcaqctqtcaatcaccaaaqtcqaqaaaqcctqaccaaccqqaacctacaqt 582
           || |||||||||||| |||| ||||||||||| ||||  ||||||| || ||||||||
Db     604 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 663

Qy     583 tqtattqccaaaqtcatctctqaaaacqqaqaqcttqqacttttcaccqatqttqttaaaqqt 642
           ||||||||  |  |||||| |||||| ||||| | ||||||| |||||| ||||||||||
Db     664 TGTATTGCCAGCTCATCTTTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGAC 723

Qy     643 aaaactqtqqgtqtaccaatagaqcccattattqctgcatqtqttqtqatctttctqaca 702
           |||  | | |   | ||   |||||||||| |||||||| ||||| ||| || | |||||
Db     724 AAACTTTTTCTCATGCCGGCAGAGCCCATCATTGCTGCTTGTGTCGTCGTCGTCCTGACG 783

Qy     703 ttqtgctttggactgattgctagaagaaaqaaaataatgaaq-ctctgcatgaaqgataa 761
           ||  |||| ||| || |||| || || | ||||||||||  | ||  ||| | ||
Db     784 ATGGCTTTTGCCCTGTTTTCTAGGAGGAAAAGAATAATGAAGTTTTTGTGGGAAAAAGAA 843

Qy     762 aqaccctcaqaqtqaaacaqctctatqa 789
           ||||| |   |||||||||||||| |
Db     844 TGACCCGAATTCAGAAACAGCTCTATAA 871


RESULT   4
BG970132
LOCUS       BG970132                797 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602839263F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4973751
            5', mRNA sequence.
ACCESSION   BG970132
VERSION     BG970132.1  GI:14357769
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10962  row: m  column: 16
            High quality sequence stop: 795.
FEATURES             Location/Qualifiers
```

```
     source          1. .797
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4973751"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: Not1;
                     Site_2: Sal1; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      249 a    174 c    181 g    193 t
ORIGIN


  Query Match           55.2%;  Score 435.4;  DB 10;  Length 797;
  Best Local Similarity   79.2%;  Pred. No. 2.8e-113;
  Matches  552;  Conservative    2;  Mismatches  138;  Indels    5;  Gaps    3;

Qy      57 tttatttctgccacgtgagatgacaagttctgttttaactgtgaatggtaaaactgagaa 116
           ||||| | ||||||  | ||  ||||| ||||||||||||||||||||| |||||||||
Db       1 TTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAATGGTAGAACTGAGAA 60

Qy     117 ctatatcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaacca 176
           |||||| ||||| ||| || ||| |||||||||||||| ||||||||||||||||||||
Db      61 CTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGTGCTGTTCAAAACCA 120

Qy     177 caccagagaggaagaactgctctggtaccgagaggaggggagagtggatttgaaatctgs 236
           ||||   || |||||||| |||||||| |||| || || | ||| ||| ||||| ||:
Db     121 CACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTAGATCTGAAAAATGG 180

Qy     237 aaacaamatcaattccagctctgtctgtgtctcttccatcagtgaaaatgacaacggaat 296
           ||||||:|||||   ||| ||||| ||||||||| |||||| |||| |||||| ||| |
Db     181 AAACAAAATCAACATCAGTTCTGTATGTGTCTCTCCCATCAACGAAAGTGACAATGGAGT 240

Qy     297 cagctttacctgcaggctgggagggatcagtccgtgtccgtttcggtggtgctgaatgt 356
           | |||||||||||| |||   | | |||||| | ||||||||| | || ||||||| ||
Db     241 CCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACAGTAGTGCTGAACGT 300

Qy     357 tacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggcagtaatgt 416
           ||| |||||||||| ||||||| || | |||||||| || |||||  |||  ||||
Db     301 TACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACAGCGATGT 360

Qy     417 gaagttggtttgcaatgtgaaagccaacccccaggctcaaatgatgtggtacaaaaacag 476
           |  |||||||||| |||||| ||||||| || || |||||||||||||| |||||||
Db     361 AAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATGTGGTATAAAAACAA 420

Qy     477 tagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtcttttca 536
           ||   | | | |||||||||| |||||||||||||||| |||||||| ||||||||||
Db     421 CAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACAAGGGAGTCTTTTCA 480

Qy     537 gctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaagtc 596
           | |||| |||||||||| |||  ||||||| || |||||||| |||||||| |  ||
Db     481 GTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGCTGTATTGCCAGCTC 540

Qy     597 atctctgaaaacggagagcttggactttcacctgattgttaaagataaaactgtgggtgt 656
           |||||||||| ||||| | ||||||| |||||| |||||||||  ||| | | |   |
Db     541 ATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGA-CAAAGTTTTGTCAT 599

Qy     657 accaatagagcccattattgctgcatgtgttgtga-tctttctgacattgtgctttggac 715
           ||   ||||||||| ||||||| ||||| |||| || | |||||     ||||| | |
Db     600 GCCGGCAGAGCCCATCATTGCTGCTTGTGTCGTGAGTCGTCCTGACGAATGGCTTTTGCC 659

Qy     716 ---tgattgctagaagaaagaaaatgaaagctctg 749
             |  ||    || |||  | | ||||||||||||
Db     660 CTGTTTTTCTAGGAGGAACAAGACTAATGAAGCTCTG 696


RESULT   5
AA986919
LOCUS       AA986919                 646 bp    mRNA    linear   EST 28-MAY-1998
DEFINITION  uc80b02.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:1431915 5' similar to TR:Q01974 Q01974 PROTEIN-TYROSINE
            KINASE TRANSMEMBRANE RECEPTOR ROR2 PRECURSOR ;, mRNA sequence.
ACCESSION   AA986919
VERSION     AA986919.1  GI:3168467
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:915983
            Possible reversed clone: similarity on wrong strand
            Seq primer: primer name ambiguous
            High quality sequence stop: 516.
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1431915"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      192 a    145 c    146 g    163 t
ORIGIN


  Query Match           53.9%;  Score 425.6;  DB 9;  Length 646;
  Best Local Similarity   81.9%;  Pred. No. 1.6e-110;
  Matches  488;  Conservative    2;  Mismatches  106;  Indels    0;  Gaps    0;

Qy      43 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 102
           ||||||||||  |||||| | |||||  | ||  ||||| ||||||||||||||||||
Db      50 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 109

Qy     103 ggtaaaactgagaactatatcctggatactacacctggctcccaagcatctctgatatgt 162
           |||| |||||||||||||| ||||| |||  || |||   ||||||||||||  ||||
Db     110 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 169

Qy     163 gctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggggagagtg 222
           ||||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db     170 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGACAAGATGGAATAGTA 229

Qy     223 gatttgaaaatctgsaaacaaamatcaattccagctctgtctgtgtctcttccatcagtgaa 282
```

```
          ||| |||||  ||:||||||:|||||   ||| |||||| |||||||| ||||||  |||
Db    240 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 289

Qy    283 aatgacaacggaatcagctttacctgcaggctgggggagggatcagtccgtgtccgtttcg 342
          | |||||| ||| || |||||||||||| |||    | | |||||| | |||||||| | |
Db    290 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 349

Qy    343 gtggtgctgaaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaag 402
          || |||||||| ||||| |||||||||||| |||||||| || | |||||||| || |||
Db    350 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 409

Qy    403 gaaggcagtaatgtgaagttggtttgcaatgtgaaaagccaaccccccaggctcaaatgatg 462
          |||  |||  |||| |  |||||||||||| |||||| ||||||| || || |||||||||
Db    410 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 469

Qy    463 tggtacaaaaacagtagtctcctcgatttagaggaaaagccgtcaccaaatccaacagaca 522
          ||||| ||||||| ||   | | | |||||||||| |||||||||||||||| ||||||
Db    470 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACA 529

Qy    523 agtgaagtcttttcagctgtcaatcaccaaaagtcgagaagcctgacaacggaacctacagt 582
          || |||||||||||| |||| |||||||||||| ||||  ||||||| || ||||||||
Db    530 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 589

Qy    583 tgtattgcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaa 638
          |||||||| | |||||||||||||| ||||| | ||||||| |||||| |||||||
Db    590 TGTATTGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAA 645


RESULT   6
BF785067
LOCUS       BF785067                 964 bp    mRNA    linear   EST 12 JAN 2001
DEFINITION  602108226F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4236606
            5', mRNA sequence.
ACCESSION   BF785067
VERSION     BF785067.1  GI:12998890
KEYWORDS    EST.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 964)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9845  row: e  column: 07
            High quality sequence stop: 761.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4236606"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      279 a    201 c    246 g    237 t      1 others
ORIGIN


  Query Match           52.7%;  Score 415.8;  DB 10;  Length 964;
  Best Local Similarity  78.2%;  Pred. No. 1.1e-107;
```

```
  Matches  559;  Conservative    2;  Mismatches  149;  Indels    5;  Gaps    5;

Qy     43 cttctcttagtaattttatttctgccacgtgagatgacaaagttctgttttaactgtgaat 102
          ||||||||||  |||||| | ||||||  | ||  |||||  ||||||||||||||||||
Db     35 CTTCTCTTAGTGGTTTTATCT-TGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 93

Qy    103 ggtaaaactgaagaactatatcctggatactacacctggctcccaagcatctctgatatgt 162
          |||| ||||||||||||||| ||||| ||| || |||  ||||||||||||  ||||
Db     94 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 153

Qy    163 gctgttcaaaaccacaccgaggatgaagaacttctctggtacagagaagatggaatagta 222
          ||||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db    154 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 213

Qy    223 gatttgaaatctgsaaacaamatcaattccagctctgtctgtgtatcttt-atcagtgaa 282
          ||| |||||  ||:||||||:|||||   ||| |||||| |||||||| ||||||  |||
Db    214 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGT-TGTGTCTCTCCCATCAACGAA 272

Qy    283 aatgacaacggaatcagctttacctgcaggctgggggagggatcagtccgtgtccgtttcg 342
          | |||||| ||| || |||||||||||| |||    | | |||||| | |||||||| | |
Db    273 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 332

Qy    343 gtggtgctgaaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaag 402
          || |||||||| ||||| |||||||||||| |||||||| || | |||||||| || |||
Db    333 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 392

Qy    403 gaaggcagtaatgtgaagttggtttgcaatgtgaaaagccaaccccccaggctcaaatgatg 462
          |||  |||  |||| |  |||||||||||| |||||| ||||||| || || |||||||||
Db    393 GAAAACAGCGATGT-AAGTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 451

Qy    463 tggtacaaaaacagtagtctcctcgatttagaggaaaagccgtcaccaaatccaacagaca 522
          ||||| ||||||| ||   | | | |||||||||| |||||||||||||||| ||||||
Db    452 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACA 511

Qy    523 agtgaagtcttttcagctgtcaatcaccaaaagtcgagaagcctgacaacggaacctacag 581
          || |||||||||||| |||| |||||| |||||| |||| ||||||| || ||||||||
Db    512 AGGGAGTCTTTTCAGTTGTCCATCACCAAAAGTCAAGAAATCTGACAATGGGACCTACAG 571

Qy    582 ttgtattgcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaaga 641
          |||||||  | ||||||||||||| |||| | ||||||| |||||| ||||||||||
Db    572 CTGTATTGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGA 631

Qy    642 taaaactgtgggtgtaccaatagagccattattgctgcatgtgttgtgatctttctgac 701
          ||  | | |   | ||  |||||||| |||||||| ||||| |||     ||
Db    632 CCAAGTTTTTGTCATGCCGGCAGAGCCCATCATTGCTGCTTGTGTCGTGGGTCGTC-CTG 690

Qy    702 attgtgctttggactgattgctagaagaaagaaaataatgaagctctgcatgaag 756
          |   || ||||| ||| || |||| || || | |  |  |  || ||    |||
Db    691 ACGATGGTTTGGCCTGTTTTCTAGGAGGAAAAGATAATGAAGTCTGTGGGAAAAG 745


RESULT   7
BM445724
LOCUS       BM445724                 538 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  11L15A10.ab1 Bos taurus Ileum #1 library Bos taurus cDNA, mRNA
            sequence.
ACCESSION   BM445724
VERSION     BM445724.1  GI:18529880
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon
            ,P.M.K. and Moore,S.S.
  TITLE     Gene Expression Profiling of the Bovine Gastrointestinal Tract
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Stephen Moore
            . Beef Genomics Laboratory
            Dept of AFNS, University of Alberta
            410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
```

```
           Tel: 780 492 0169
           Fax: 780 492 4265
           Email: smoore@afns.ualberta.ca
           Insert Length: 538   Std Error: 0.00
           POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="Bos taurus Ileum #1 library"
                     /tissue_type="Smooth muscle"
                     /cell_type="Simple columnar epithelial"
                     /dev_stage="Young adult"
                     /lab_host="XL1-BlueMRF'strain"
                     /note="Organ: Intestine/ileum; Vector: Uni-ZAPXR; Site_1:
                     EcoRI; Site_2: Xho I"
BASE COUNT      165 a    118 c    121 g    134 t
ORIGIN
```

```
  Query Match             51.0%;  Score 402.6;  DB 10;  Length 538;
  Best Local Similarity   85.7%;  Pred. No. 5.8e-104;
  Matches  457;  Conservative    2;  Mismatches   71;  Indels    3;  Gaps    1;

Qy      66 gccacgtgagatgacaagttctgttttaactgtgaatggtaaaactgagaactatatcct 125
           ||||| || |||||||||||||||||||| |||||||||||||||||| |||||||| ||
Db       5 GCCACATGTGATGACAAGTTCTGTTTTAAGTGTGAATGGTAAAACTGAAAACTATATTCT 64

Qy     126 ggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacaccagaga 185
           ||| |||   |||||  ||||| |||||||| ||||||||||||||||||| | ||||
Db      65 GGACACTGAGCCTGGCCTCCAAGAATCTCTGAAATGTGCTGTTCAAAACCACATCCGAGA 124

Qy     186 ggaagaactgctctggtaccgagaggaggggagagtggatttgaaatctgsaaacaamat 245
           |||||||| |||||||||| || || |||||||||||||||||| ||||:|||||:||
Db     125 TGAAGAACTTCTCTCGTACCGCGAAGATGGGAGAGTGGATTTGAAGTCTCGAAACAAAAT 184

Qy     246 caattccagctctgtctgtgtctcttccatcagtgaaaatgacaacggaatcagctttac 305
           ||| |||||||||||||||||||    |||||||||| |||||||||| |||| ||||||
Db     185 CAACTCCAGCTCTGTCTGTGTCTCGGGCATCAGTGAAGATGACAACGGCATCACCTTTAC 244

Qy     306 ctgcaggctggggagggatcagtccgtgtccgtttcggtggtgctgaatgttacttttcc 365
           ||||| ||||  |||| |||||||| |||||| | || |||||||||||| || ||||||||
Db     245 CTGCAAGCTGCAGAGGAATCAGTCGGTGTCCATCTCAGTGGTGCTGAACGTCACTTTTCC 304

Qy     366 tcctctcctaagtggaaacgacttccaaacagttgaggaaggcagtaatgtgaagttggt 425
           |||||||||||||||||| |||||||||||||| |||||||||||||| ||||||||||||
Db     305 TCCTCTCCTAAGTGGAAATGACTTCCAAACAGCTGAGGAAGGCAGTGATGTGAAGTTGGT 364

Qy     426 ttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtacaaaaacagtagtctcct 485
           |||||||||||| |||||||||||||||||| |||||||||| ||||||| | |  || |
Db     365 TTGCAATGTGAAATCCAACCCCCAGGCTCAGATGATGTGGTATAAAAACAATGGCATCTT 424

Qy     486 cgatttagagaaaagccgtcaccaaatccaacagacaagtgagtcttttcagctgtcaat 545
           | ||||||||   || ||||||||||||||||||||  |||  |||||  |||| ||
Db     425 GAACTTAGAGAA---CCATCACCAAATCCAACAGACAAGCAAGTACTTTCAATTGTCGAT 481

Qy     546 caccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaagtcat 598
           |||||||||| ||||  | |||| || || || || |||||||| || ||||
Db     482 CACCAAAGTCAAGAAATCCGACAATGGGACTTATAGCTGTATTGCCAATTCAT 534
```

```
RESULT   8
BF788117
LOCUS       BF788117                 936 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  602133341F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4241194
            5', mRNA sequence.
ACCESSION   BF788117
VERSION     BF788117.1  GI:12093166
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 936)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9857  row: b  column: 11
            High quality sequence start: 5
            High quality sequence stop: 713.
FEATURES             Location/Qualifiers
     source          1. .936
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4241194"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      270 a    213 c    231 g    220 t      2 others
ORIGIN
```

```
  Query Match             48.4%;  Score 382.2;  DB 10;  Length 936;
  Best Local Similarity   80.8%;  Pred. No. 4.4e-98;
  Matches  492;  Conservative    2;  Mismatches  110;  Indels    5;  Gaps    4;

Qy      43 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 102
           ||||||||||  |||||| | ||||||  | ||  ||||| || ||||||||||||||||
Db      31 CTTCTCTTAGTGGTTTTATCT-TGCCACAGGGGAGAACAAGCTC-GTTTTAACTGTGAAT 88

Qy     103 ggtaaaactgagaactatatcctggatactacacctggctcccaagcatctctgatatgt 162
           |||| ||||||||||||||| ||||| |||  || |||   |||||||||||||  ||||
Db      89 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 148

Qy     163 gctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggggagagtg 222
           ||||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db     149 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 208

Qy     223 gatttgaaatctgsaaacaamatcaattccagctctgtctgtgtctcttccatcagtgaa 282
           ||| |||||  ||:|||||:|||||   ||| |||| |||||||| ||||||   |||
Db     209 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGT-TGTGTCTCTCCCATCAACGAA 267

Qy     283 aatgacaacggaatcagctttacctgcaggctgggggagggatcagtccgtgtccgtttcg 342
           | |||||| ||| || |||||||||||| |||   | | |||||| | ||||||||| |
Db     268 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 327

Qy     343 gtggtgctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgag 402
           || |||||||| |||| |||||||||||| |||||||| || | |||||||||| || |||
Db     328 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 387

Qy     403 gaaggcagtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatg 462
           |||  |||  ||||  |  | | |||||| |||||| |||||| || || |||||||||
Db     388 GAAAACAGCGATGT--AAGTTGGTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 445

Qy     463 tggtacaaaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagaca 522
           ||||| |||||||  ||   | | | |||||||||| ||||||||||||||||| ||||||
Db     446 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACA 505

Qy     523 agtgagtcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagt 582
           || |||||||||| |||| |||||||||||| ||||  ||||||| || ||||||||
Db     506 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 565

Qy     583 tgtattgcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagat 642
```

```
          |||||| || |  |||||||||||||| ||||| | ||||||| |||||| ||||||||||
Db    566 TGTATCGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGAC 625

Qy    643 aaaactgtg 651
          |||  | ||
Db    626 AAAGTTTTG 634


RESULT   9
AA500290
LOCUS       AA500290                 543 bp    mRNA    linear   EST 01-JUL-1997
DEFINITION  vi97q08.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:920222 5' similar to TR:G337467 G337467
            PROTEIN-TYROSINE KINASE TRANSMEMBRANE RECEPTOR ROR2 PRECURSOR ;,
            mRNA sequence.
ACCESSION   AA500290
VERSION     AA500290.1  GI:2235257
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:532438
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 431.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:920222"
                     /clone_lib="Barstead mouse pooled organs MPLRB4"
                     /sex="mixed"
                     /tissue_type="pooled organs"
                     /dev_stage="7 day"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [GTTGGATTCGGTACC], digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead."
BASE COUNT      164 a    123 c    121 g    135 t
ORIGIN


  Query Match             48.4%;  Score 481.8;  DB 9;  Length 543;
  Best Local Similarity   81.3%;  Pred. No. 5e-98;
  Matches  440;  Conservative    2;  Mismatches   99;  Indels    0;  Gaps    0;

Qy      65 tgccacgtgaqatgacaagttctgttttaactgtgaatggtagaactgagaactatattcc 124
           |||||| || ||  ||||| ||||||||||||||||||||||| ||||||||||||||| |
Db       2 TGCCACATGGGAGAACAAGCTCTGTTTTAACTGTGAATGGTAGAACTGAGAACTATATTC 61

Qy     125 tgqatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacaccaqaq 184
           |||| ||| || |||   |||||||||||||  |||||||||||||||||||||||||  |
Db      62 TGGACACTCAACATGGTGTCCAAGCATCTTGGAATGTGCTGTTCAAAACCACACCCAGG 121

Qy     185 aggaaqaactgctctggtaccqagaggaggggqaqaqtggatttgaaatctgsaaacaama 244
           | ||||| || |||||||||| |||| || || | ||| ||| |||||  ||:|||||:|
Db     122 ATGAAGATCTTCTCTGGTACAGAGAAGATGGAATAGTAGATCTGAAAAATGGAAACAAAA 181

Qy     245 tcaattccaqctctgtctgtgtctcttccatcagtgaaaatqacaacqgaatcaqcttta 304
           ||||   ||| ||||| |||||||||| ||||||  |||| |||||| ||| || ||||||
Db     182 TCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAAAGTGACAATGGAGTCCGCTTTA 241

Qy     305 cctgcaggctggggagggatcagtccgtgtccgtttcggtgtqgtgctqaatgttacttttc 364
           |||||| |||   | | |||||| | |||||||||| | || |||||||| ||||| ||||
Db     242 CCTGCAAGCTACAGCCCGATCAGACGGTGTCCGTTACAGTAGTGCTGAACGTTACGTTTC 301

Qy     365 ctcctctctctaagtggaaacgacttccaaacagttgaggaaggcagtaatgtgaagttqg 424
           ||||||| |||||||| || | |||||||||| || ||||||  |||  |||| |  ||||
Db     302 CTCCTCTTCTAAGTGGCAATGGCTTCCAAACGGTCGAGGAAAACAGCGATGTAAGTTTGG 361

Qy     425 tttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtacaaaaacagtagtctcc 484
           |||||| |||||| ||||||| || || |||||||||||||| |||||||  ||   |
Db     362 TTTGCAACGTCAAATCCAACCCTCAAGCGCAAATGATGTGGTATAAAAACAACAGCGCCT 421

Qy     485 tcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtcttttcagctgtcaa 544
           | | |||||||||| |||||||||| ||||| ||||||| |||||||||||| |||| |
Db     422 TGGTTTTAGAGAAAGGCCGTCACCACATCCACCAGACAAGGGAGTCTTTTCAGTTGTCCA 481

Qy     545 tcaccaaagtcgagaagcctgacaacqgaacctacagttgtattgccaaagtcatctctga 604
           |||||||||| |||| |||||| || ||||||| ||||||| |  ||||||||||
Db     482 TCACCAAAGTCAAGAAATCTGACATTGGGACCTACAGCTGTATTGCCAGCTCATCTCTGA 541

Qy     605 a 605
           |
Db     542 A 542


RESULT  10
AI747948
LOCUS       AI747948                 657 bp    mRNA    linear   EST 22-JUN-1999
DEFINITION  ul04c10.y1 Sugano Mouse Kidney mkia Mus musculus cDNA clone
            IMAGE:2065074 5' similar to TR:Q01974 Q01974 PROTEIN-TYROSINE
            KINASE TRANSMEMBRANE RECEPTOR ROR2 PRECURSOR ;, mRNA sequence.
ACCESSION   AI747948
VERSION     AI747948.1  GI:5126212
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:994262
            Seq primer: custom primer used
            High quality sequence stop: 520.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Mus musculus"
```

```
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2065074"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert   Size selection was
                     performed to exclude fragments <1.5kb   Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      195 a    151 c    153 g    158 t
ORIGIN


  Query Match             42.6%;  Score 336.2;  DB 9;  Length 657;
  Best Local Similarity   78.8%;  Pred. No. 5.5e-85;
  Matches  401;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy     281 aaaatgacaacggaatcagctttacctgcaggctggggagggatcagtccgtgtccgtttt 340
           |||||||| ||| || ||||||||||||| |||   | | |||||| | ||||||||||
Db      82 ACAATGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTA 141

Qy     341 cggtggtgctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttg 400
           | || |||||||| ||||| |||||||||||| |||||||| || | |||||||||| || |
Db     142 CAGTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGG 201

Qy     401 aggaaggcagtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatga 460
           ||||| ||| |||| | |||||||||| |||||| ||||||| || || |||||||
Db     202 AGGAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGA 261

Qy     461 tgtggtacaaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacaga 520
           ||||||| ||||||| ||   | | | |||||||||| ||||||||||||||||| ||||
Db     262 TGTGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGA 321

Qy     521 caagtgagtcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctaca 580
           |||| |||||||||||| |||| |||||||||||| ||||  ||||||| || |||||||
Db     322 CAAGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACA 381

Qy     581 gttgtattgcaaagtcatctctgaaaacggagagcttggacttt-aaccgattgttaaag 640
           | |||||||| |  |||||||||||| ||||| | |||||| |||||| |||||||||
Db     382 GCTGTATTGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAG 441

Qy     641 ataaaactgtgggtgtaccaatagagcccattattgctgcatgtgttgtgatctttctga 700
           | |||  | | |   | ||   |||||||| |||||||| ||||| ||| || | ||||
Db     442 ACAAAGTTTTTGTCATGCCGGCAGAGCCCATCATTGCTGCTTGTGTCGTGGTCGTCCTGA 501

Qy     701 cattgtgctttggactgattgctagaagaaagaaaataatgaagctctgcatgaaggata 760
           |  ||   |||| ||| || |||| || || | ||||||||||||||   ||| | |
Db     502 CGATGGCTTTTGCCCTGTTTTCTAGGAGGAAAAGAATAATGAAGCTCTGTGGGAAAAAGA 561

Qy     761 aagaccctcacagtgaaacagctctatga 789
           | |||||  |    ||||||||||||| |
Db     562 ATGACCCGAATTCAGAAACAGCTCTATAA 590


RESULT  11
AI956638
LOCUS       AI956638                 445 bp    mRNA    linear   EST 20-AUG-1999
DEFINITION  ul79f12 y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2136815 5' similar to SW:VGR2_HUMAN P35968 VASCULAR
            ENDOTHELIAL GROWTH FACTOR RECEPTOR 2 PRECURSOR ;, mRNA sequence.
ACCESSION   AI956638
VERSION     AI956638.1  GI:5749347
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Papé,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St  Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1001491
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2136815"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      136 a     94 c    102 g    113 t
ORIGIN


  Query Match             38.3%;  Score 306.6;  DB 9;  Length 445;
  Best Local Similarity   80.2%;  Pred. No. 1.4e-76;
  Matches  357;  Conservative    2;  Mismatches   86;  Indels    0;  Gaps    0;

Qy      57 tttatttctgccacgtgagatgacaagttctgttttaactgtgaatggtaaaactgagaa 116
           ||||| | |||||  | ||  |||| ||||||||||||||||||||||| ||||||||
Db       1 TTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAATGGTAGAACTGAGAA 60

Qy     117 ctatatcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaacca 176
           |||||| |||| ||| || |||   |||||||||||||  ||||||||||||||||||
Db      61 CTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGTGCTGTTCAAAACCA 120

Qy     177 caccagagaggaagaactgctctggtaccgagaggaggggagagtggatttgaaatctgg 236
           ||||   || |||||||| ||||||||| |||| || || | ||| ||| |||||  ||:
Db     121 CACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTAGATCTGAAAAATGG 180

Qy     237 aaacaaatcaattccagctctgtctgtgtctcttccatcagtgaaaatgacaacggaat 296
           ||||||:|||||   ||| |||| |||||||| ||||||  |||| |||||| ||| |
Db     181 AAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAAAGTGACAATGGAGT 240

Qy     297 cagctttaccgcaggctggggagggatcagtccgtgtccgtttcggtggtgctgaatgt 356
```

```
          | ||||||||||||| |||   | | |||||| | |||||||||| | || ||||||||| ||
Db     241 CCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACAGTAGTGCTGAACGT 300

Qy     357 tacttttcctcctctcctaaqtqqaaacqactccaaacaqttqaggaaggcaqtaatqt 416
           ||| |||||||||| ||||||||| || | |||||||||| || ||||||  |||  ||||
Db     301 TACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACAGCGATGT 360

Qy     417 qaaqttqqtttqcaatqtqaaaqccaacccccaqqctcaaaatqatqtqqtacaaaaaacaq 476
           |  |||||||||||| |||||| ||||||| || || ||||||||||||||| |||||||
Db     361 AAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATGTGGTATAAAAACAA 420

Qy     477 taqtctcctcqatttaqaqaaaaqc 501
           ||   | | | ||||||||||| ||
Db     421 CAGCGCCTTGGTTTTAGAGAAAGGC 445


RESULT  12
AI615364
LOCUS       AI615364                 487 bp    mRNA    linear   EST 21-Apr-1999
DEFINITION  vi97q08.y1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:920222 5', mRNA sequence.
ACCESSION   AI615364
VERSION     AI615364.1  GI:4624531
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:532438
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 478.
FEATURES             Location/Qualifiers
     source          1..487
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:920222"
                     /clone_lib="Barstead mouse pooled organs MPLRB4"
                     /sex="mixed"
                     /tissue_type="pooled organs"
                     /dev_stage="7 day"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D Pac (Pharmacia) with
                     a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [GTTGGATTCGGTACC], digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead."
BASE COUNT      146 a    139 c    111 g    120 t      1 others
ORIGIN
```

```
  Query Match           37.5%;  Score 295.6;  DB 9;  Length 487;
  Best Local Similarity   80.8%;  Pred. No. 2e-73;
  Matches  354;  Conservative    2;  Mismatches   81;  Indels    1;  Gaps    1;

Qy      43 cttctcttagtggttttatctttgccacaggggagaacaagctctgttttaactgtgaat 102
           ||||||||||  |||||| | ||||||  | ||  ||||| ||||||||||||||||||
Db      47 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 106

Qy     103 qqtaaaactqaqaactatatcctqqatactacacctqqctcccaaqcatctctqatatqt 162
           |||| |||||||||||||||| ||||| |||  || |||   |||||||||||||  ||||
Db     107 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 166

Qy     163 qctqttcaaaaccacaccgaggatgaagaacttctctggtacagagaagatggaatagta 222
           ||||||||||||||||||    || |||||||| |||||||| |||| || || | |||
Db     167 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 226

Qy     223 qatttqaaatctqsaaacaamatcaattccaqctctqtctqtctcttccatcaqtqaa 282
           ||| |||||  ||:|||||:|||||   ||| ||||| |||||||| |||||| |||
Db     227 CATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCCCATCAACCAA 286

Qy     283 aatqacaacqgaatcaqctttaccqtqcaggctqqqqaqggatcaqtccqtqtccqtttcq 342
           | |||||| ||| || |||||||||||| |||   | | |||||| | |||||||| |
Db     287 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 346

Qy     343 qtqqtqctqaatqttacttttcctcctctcctaaqtqqaaacqacttccaaacaqttqaq 402
           || |||||||| ||||| |||||||||| |||||||| || | |||||||| || |||
Db     347 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 406

Qy     403 qaaqqcaqtaatqtqaaqttqqtttqcaatqtqaaaqccaacccccaqqctcaaatqa-t 461
           |||  |||  |||| |  |||||||||| |||||| ||||||| || || ||||||| |
Db     407 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGANT 466

Qy     462 qtqqtacaaaaacaqtaq 479
           |||||| |||||||  ||
Db     467 GTGGTATAAAAACAACAG 484


RESULT  13
AW211197
LOCUS       AW211197                 514 bp    mRNA    linear   EST 03-DEC-1999
DEFINITION  ul25h03.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2099381 5' similar to TR:O60469 O60469 DOWN SYNDROME CELL
            ADHESION MOLECULE. ;, mRNA sequence.
ACCESSION   AW211197
VERSION     AW211197.1  GI:6517137
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: ul25h03.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:996313
            Seq primer: custom primer used
            High quality sequence stop: 513.
FEATURES             Location/Qualifiers
     source          1..514
                     /organism="Mus musculus"
```

/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2099381"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing. 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."
BASE COUNT 150 a 112 c 123 g 129 t
ORIGIN

Query Match 30.6%; Score 241.2; DB 9; Length 514;
Best Local Similarity 78.3%; Pred. No. 6.7e-58;
Matches 299; Conservative 2; Mismatches 80; Indels 1; Gaps 1;

```
Qy      43 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 102
           ||||||||||| |||||| | |||||| | ||  ||||| ||||||||||||||||| |
Db     133 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAGT 192

Qy     103 ggtaaaactgagaactatatcctggatactacacctggctcccaagcatctctgatatgt 162
           |||| ||||||||||||||| ||||| ||| || |||   ||||||||||||  ||||
Db     193 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 252

Qy     163 gctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggggagagtg 222
           ||||||||||||||||||   || |||||||| ||||||||| |||| || || | |||
Db     253 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 312

Qy     223 gatttgaaatctgsaaacaamatcaattccagctctgtctgtgtctcttccatcagtgaa 282
           ||| ||||| ||:||||||:|||||   ||| ||||| |||||||| |||||| |||
Db     313 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 372

Qy     283 aatgacaacggaatcagctttacctgcaggct-ggggagggatcagtccgtgtccgtttc 341
           | |||||| ||| || |||||||||||| |||   | | |||||| |||||||| | |
Db     373 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACCAGCGCGATCAGACCGTGTCCCGTAC 432

Qy     342 ggtggtgctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttga 401
           || |||||  |  ||| |||||||||||| |||||||| || | ||| ||||| || ||
Db     433 AGTAGTGCTTGACCGTACCTTTCGTCCTCTTCTAAGTGGCAATGGCTTTCAAACCGTGGA 492

Qy     402 ggaaggcagtaatgtgaagttg 423
           ||||  |||  |||| |  |||
Db     493 GGAAAACAGCGATGTAAGTTTG 514
```

RESULT 14
BB845854
LOCUS BB845854 437 bp mRNA linear EST 26-NOV-2001
DEFINITION BB845854 RIKEN full-length enriched, adult male kidney Mus musculus cDNA clone F520008G20 5', mRNA sequence.
ACCESSION BB845854
VERSION BB845854.1 GI:17084229
KEYWORDS EST.
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 437)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http.//genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .437
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="F520008G20"
/clone_lib="RIKEN full-length enriched, adult male kidney"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAAGTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. "
BASE COUNT 125 a 99 c 99 g 114 t
ORIGIN

Query Match 27.3%; Score 215.4; DB 9; Length 437;
Best Local Similarity 79.8%; Pred. No. 1.5e-50;
Matches 264; Conservative 2; Mismatches 63; Indels 2; Gaps 1;

```
Qy      43 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 102
           ||||||||||| |||||| | |||||| | ||  ||||| ||||||||||||||||| |
Db     107 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 166
```

```
Qy      103 qqtaaaactqaqaactatatcctqqatactacacctqqctcccaaqcatctctqatatqt 162
            |||| ||||||||||||||| ||||| ||| || |||   ||||||||||||  ||||
Db      167 GGTACAACTCACAACTATATTCTCCACACTCAACATCCTCTCCAAGCATCTCTCGGAATGT 226

Qy      163 ptptttcaaaaccacaccqacqatqaaqaactctctqqtacaqaqaaggaqtttatqatqtq 222
            ||||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db      227 GCTGTTCAAAACCACACCGACGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 286

Qy      223 gatttqaaaatctqsaaacaaatqatcaattccaqctctqtctqtqtctcttccatcaqtqaa 282
            ||| |||||  ||:|||||:|||||   ||| ||||| |||||||| |||||  |||
Db      287 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 346

Qy      283 aatqaccaacqqaatcaqc--tttacctqcaqqctqqqqaqqqatcaqtccqtqtccqttt 340
            | |||||| ||| || ||  ||  | |||| |||   | | |||||| | |||||||||
Db      347 AGTGACAATGGAGTCCGCCTTTACCGTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTA 406

Qy      341 cqqtqqtqctqaatqttacttttcctcctct 371
            | || |||||||| ||||| ||||||||||
Db      407 CAGTAGTGCTGAACGTTACCTTTCCTCCTCT 437


RESULT  15
AZ864346/c
LOCUS       AZ864346               515 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0173F20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0173F20 R, DNA sequence.
ACCESSION   AZ864346
VERSION     AZ864346.1  GI:13063555
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,K., Tingey,A., Von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0173  row: F  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 515.
FEATURES             Location/Qualifiers
     source          1..515
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0173F20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD4Znv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      138 a    117 c    105 g    155 t
ORIGIN


  Query Match             24.1%;  Score 189.8;  DB 12;  Length 515;
  Best Local Similarity   81.0%;  Pred. No. 3.2e-43;
  Matches  221;  Conservative    0,  Mismatches   52;  Indels    0;  Gaps    0;

Qy      378 tggaaacgacttccaaacagttgaqqaaqqcagtaatgtgaaqttgqtttqcaatgtqaa 437
            ||| || | |||||||||| || ||||||  |||  |||| |  ||||||||||| |||||
Db      515 TCGCAATGGCTTCCAAACCGTGGAGGAAAACAGCGATGTAAGTTTCGTTTGCAACGTGAA 456

Qy      438 agccaacccccaqqctcaaatqatqtqqtacaaaaacaqtaqtctcctcqatttaqaqaa 497
            | ||||||| || || |||||||||||||| |||||||  ||     | | |||||||||
Db      455 ATCCAACCCTCAAGCCCAAATGATGTGGTATAAAAACAACAGCGCTTTGGTTTTAGAGAA 396

Qy      498 aaqccqtcaccaaatccaacaqacaaqtqaqtctttcaqctqtcaatcaccaaaqtcqa 557
            | |||||||||||||||| |||||||| |||||||||||| |||| ||||||||||| |
Db      395 AGGCCGTCACCAAATCCACCAGACAAGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAA 336

Qy      558 qaaqcctqacaacqqaacctacaqttqtattqcaaaqtcatctctqaaaacqqaqaqctt 617
            |||  ||||||| || || |||| |||||||| |  ||||||||||||| |||| | |
Db      335 GAAATCTGACAATGGCACTTACAGCTGTATTGCCAGCTCATCTCTCAAAATGGAGACCAT 276

Qy      618 qqactttcaccqattqttaaaqqtaaaactqt 650
            |||||| ||| || |||||||| |||  || |
Db      275 GCACTTCCACTTGCTTGTTAAAGGTAACTCTCT 243


Search completed. September 17, 2002, 09:04:22
Job time: 2575 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 08:21:27 ; Search time 2548.21 Seconds
(without alignments)
4051.931 Million cell updates/sec

Title: US-09-863-823-1
Perfect score: 765
Sequence: 1 atgcaaatgggaagatttct.....acagtgaaacagctctatga 765

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	655.8	85.7	822	10	BI764721	BI764721 603051058
2	504.6	66.0	1367	11	AK009118	AK009118 Mus muscu
3	487.2	63.7	1317	11	AK008028	AK008028 Mus muscu
4	435.4	56.9	797	10	BG970132	BG970132 602839263
5	425.6	55.6	646	9	AA986919	AA986919 uc80b02.y
6	415.8	54.4	964	10	BF785067	BF785067 602108226
7	402.6	52.6	538	10	BM445724	BM445724 11L15A10.
8	382.2	50.0	936	10	BF788117	BF788117 602113341
9	381.8	49.9	543	9	AA500290	AA500290 vi97g08.r
10	336.2	43.9	657	9	AI747948	AI747948 ul04c10.y
11	306.6	40.1	445	9	AI956638	AI956638 ul79f12.y
12	295.6	38.6	487	9	AI615364	AI615364 vi97g08.y
13	241.2	31.5	514	9	AW211197	AW211197 ul25h03.y
14	215.4	28.2	437	9	BB845854	BB845854 BB845854
c 15	189.8	24.8	515	12	AZ864346	AZ864346 2M0173F20
c 16	160	20.9	504	9	AW235098	AW235098 xn18e10.x
c 17	156.8	20.5	175	9	AW844313	AW844313 RC2-CN005
18	131.8	17.2	535	9	AW107007	AW107007 um18d09.y
19	108	14.1	504	10	W82090	W82090 me98e12.r1
20	106.4	13.9	320	9	AI425843	AI425843 me98e12.y
21	89.8	11.7	125	10	BM445552	BM445552 11L12H10.
22	61.2	8.0	545	10	BM026157	BM026157 fu75c12.y
c 23	52	6.8	572	9	AI790737	AI790737 ul04c10.x
c 24	51.4	6.7	590	10	BM025920	BM025920 fu75c12.x
25	43.8	5.7	151	9	AA489845	AA489845 ab04g06.s
c 26	43.4	5.7	522	9	AW107893	AW107893 um18d09.x
27	40.6	5.3	432	9	AA772324	AA772324 ai43a08.s
28	40.6	5.3	780	12	CN504142	AL291788 Tetraodon
c 29	40.2	5.3	393	10	W94404	W94404 zd80f12.s1
30	40.2	5.3	411	12	AZ979833	AZ979833 2M0256107
31	39.8	5.2	435	9	AI672771	AI672771 we58f05.x
32	39.2	5.1	433	9	AI032799	AI032799 ow17c10.x
33	39.2	5.1	513	9	AI160022	AI160022 qc91b04.x
c 34	39.2	5.1	523	10	BF063348	BF063348 7h88e07.x
35	39.2	5.1	637	10	BF350315	BF350315 hf13b10.x
36	39.2	5.1	670	10	BM171749	BM171749 imagegc_5
37	39.2	5.1	816	9	AU117395	AU117395 AU117395
38	39	5.1	252	10	BF785567	BF785567 602112349
39	39	5.1	474	9	BB818563	BB818563 BB818563
c 40	39	5.1	610	12	FR002997	AL014866 F.rubripe
41	39	5.1	642	9	BB652201	BB652201 BB652201
42	39	5.1	665	9	BB614047	BB614047 BB614047
43	39	5.1	676	9	BB465902	BB465902 BB465902
44	39	5.1	940	11	AK020591	AK020591 Mus muscu
45	38.2	5.0	478	9	AA844204	AA844204 ai91h02.s

ALIGNMENTS

RESULT 1
BI764721
LOCUS BI764721 822 bp mRNA linear EST 25-SEP-2001
DEFINITION 603051058F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5190878 5', mRNA sequence.
ACCESSION BI764721
VERSION BI764721.1 GI:15756299
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11477 row: d column: 15
High quality sequence stop: 821.
FEATURES Location/Qualifiers
source 1..822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5190878"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.4 kb, insert size range 1 3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."
BASE COUNT 251 a 166 c 195 g 210 t
ORIGIN

Query Match 85.7%; Score 655.8; DB 10; Length 822;
Best Local Similarity 97.5%; Pred. No. 2.4e-177;
Matches 738; Conservative 2; Mismatches 9; Indels 8; Gaps 7;

```
Qy       1 atgcaaatgtggaagatttcttctcttagtaatttttatttctgccacgtgagatgacaagt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17 ATGCAAATGGGAAGATTTCTTCTCTTAGTAATTTTTATTTCTGCCACGTGAGATGACAAGT 76

Qy      61 tctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacctggctcc 120
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 TC-GTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCTGGCTCC 135

Qy     121 caagcatctctgatatgtgctgttcaaaaccacaccagaagaggaaagaactgctctggtac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 CAAGCATCTCTGATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTCTGGTAC 195

Qy     181 cgagaggagggagagtggatttgaaatctgsaaacaamatcaattccagctctgtctgt 240
           |||||||||||||||||| ||||||||||:||||||:|||||||||||||||||||||||
Db     196 CGAGAGGAGGGAGAGTGGA-TTGAAATCTGGAAACAAAATCAATTCCAGCTCTGTCTGT 254

Qy     241 gtctcttccatcagtgaaaatgacaacggaatcagctttacctgcaggctgggagggat 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     255 GTCTCTTCCATCAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGGAGGGAT 314

Qy     301 cagtcccgtgtcccgtttcggtggtgctgaatgttacttttcctcctctcctaagtggaaac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     315 CAGTCCGTGTCCGTTTCGGTGGTGCTGAATGTTACTTTTCCTCCTCTCCTAAGTGGAAAC 374

Qy     361 gacttccaaacagttgaggaaggcagtaatgtgaagttggtttgcaatgtgaaagccaac 420
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     375 GACTTCCAAACAGTTGAGGAAGGCAGTAATGTGAAG-TGGTTTGCAATGTGAAAGCCAAC 433

Qy     421 ccccagqctcaaatgatgtggtacaaaaacagtagtctcctcgatttagagaaaagccgt 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     434 CCCCAGGCTCAAATGATGTGGTACAAAAACAGTAGTCTCCTCGATTTAGAGAAAAGCCGT 493

Qy     481 caccaaatccaacagacaagtgagtcttttcagctgtcaatcaccaaagtcgagaagcct 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     494 CACCAAATCCAACAGACAAGTGAGTCTTTTCAGCTGTCAATCACCAAAGTCGAGAAGCCT 553

Qy     541 gacaacggaacctacagttgtattgcaaaagtcatctctgaaaacggagagacttggacttt 600
           ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     554 GACAACGGAACCTACAGTTGTATGGCAAAGTCATCTCTGAAAACGGAGAGCTTGGACTTT 613

Qy     601 cacctgattgttaaagataaaactgtgggtgtaccaatagagcccattattgctgcatgt 660
           ||||||| ||||||||||||||||||||||||||||||||||||||||| |||||||
Db     614 CACCTGA-TGTTAAAGATAAAACTGTGGGTGTACCAATAGAGCCCATTATTG-TGCATGT 671

Qy     661 gttgtgatctttctgacattgtgcttt-ggactgattgctagaagaaagaaaatagaag 720
           |||||||||||||||||||||||||| | |||||||||||||||| ||||||||
Db     672 GTTGTGATCTTTCTGACATTGTGCTT--GGACTGATGCTAGAAGAAAGAACATAATGAAG 729

Qy     721 ctctgcatgaaggatagagaccctcacagtgaaacag 757
           |||||| |||||||||||| ||||||||||||||||
Db     730 CTCTGCTTGAAGGATAAAGA-CCTCACAGTGAAACAG 765
```

RESULT 2
AK009118
LOCUS AK009118 1367 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full length enriched library, clone:2310003K08-Immunoglobulin domain containing protein, full insert sequence.
ACCESSION AK009118
VERSION AK009118.1 GI:12843711
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full length enriched mouse cDNA library clone:2310003K08.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1367)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,I., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full length by cap-trapper. cDNA went

through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

```
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1898129"
                     /db_xref="taxon:10090"
                     /clone="2310003K08"
                     /sex="male"
                     /tissue_type="tongue"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             24. .809
                     /note="Immunoglobulin domain containing protein
                     data source:Pfam, source key:PF00047, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB26083.1"
                     /db_xref="GI:12843712"
                     /translation="MVWKITGPLQACQLLLVVLSLPQGRTSSVLTVNGRTENYILDTQ
                     HGVQASLECAVQNHTEDEELLWYREDGIVDLKNGNKINISSVCVSPINESDNGVRFTC
                     KLQRDQTVSVTVVLNVTFPPLLSGNGFQTVEENSDVSLVCNVKSNPQAQMMWYKNNSA
                     LVLEKGRHQIHQTRESFQLSITKVKKSDNGTYSCIASSSLKMETMDFHLLVKDKVFVM
                     PAEPIIAACVVVVLTMAFALFSRRKRIMKLCGKKNDPNSETAL"
BASE COUNT       417 a    267 c    298 g    385 t
ORIGIN
```

Query Match 66.0%; Score 504.6; DB 11; Length 1367;
Best Local Similarity 79.5%; Pred. No. 7.3e-134;
Matches 594; Conservative 2; Mismatches 151; Indels 0; Gaps 0;

```
Qy      19 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 78
           ||||||||||  |||||| | ||||||  | ||  |||||  ||||||||||||||||||
Db      63 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 122

Qy      79 ggtaaaactgagaactatatcctggatactacacctggctcccaagcatctctgatatgt 138
           ||||  ||||||||||||||| |||||  |||  ||  |||   ||||||||||||  ||||
Db     123 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 182

Qy     139 gctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggggagagtg 198
           ||||||||||||||||||   || ||||||||  |||||||| |||| || || | |||
Db     183 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGGAATAGTA 242

Qy     199 gatttgaaatctgsaaacaamatcaattccagctctgtctgtgtctcttccatcagtgaa 258
           |||  |||||  ||:||||||:|||||   ||| |||||  |||||||| ||||||  |||
Db     243 GATCTCAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 302

Qy     259 aatgacaacggaatcagctttacctgcaggctggggagggatcagtccgtgtccgtttcg 318
           | |||||| ||| || ||||||||||||| |||   | | |||||| | |||||||||| |
Db     303 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 362

Qy     319 gtggtgctgaatgttacttttcctcttctcctaagtggaaacgactttcaaacagttgag 378
           || |||||||| |||| ||||||||||| |||||||| || | ||||||||| || |||
Db     363 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 422

Qy     379 gaaggcagtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatg 438
           |||  |||  |||| |  |||||||||| |||||| ||||||| || || |||||||||
Db     423 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 482

Qy     439 tggtacaaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagaca 498
           |||||  |||||||  ||   | | | ||||||||||  |||||||||||||||| ||||||
Db     483 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACA 542

Qy     499 agtgagtcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagt 558
           || |||||||||| |||| |||||||||||| ||||  ||||||| || |||||||||
Db     543 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 602

Qy     559 tgtattgcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagat 618
           |||||||| |  |||||||||||||| ||||| | ||||||| |||||| ||||||||||
Db     603 TGTATTGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGAC 662

Qy     619 aaaactgtgggtgtaccaatagagcccattattgctgcatgtgttgtgatctttctgaca 678
           |||  | | |   | ||  |||||||||| |||||||| ||||| ||| || | |||||
Db     663 AAAGTTTTTGTCATGCCGGCAGAGCCCATCATTGCTGCTTGTGTCGTGGTCGTCCTGACG 722

Qy     679 ttgtgctttggactgattgctagaagaaagaaaataatgaaagctctgcatgaaggataaa 738
           ||   ||||  ||| || |||| || || | ||||||||||||||||   ||| | ||
Db     723 ATGGCTTTTGGCCTGTTTTCTAGGAGGAAAAGAATAATGAAGCTCTGCATGAAAAAGAAT 782

Qy     739 gacccctcacagtgaaacagctctatga 765
           |||||  |   ||||||||||||| |
Db     783 GACCCGAATTCAGAAACAGCTCTATAA 809
```

```
RESULT   3
AK008028
LOCUS       AK008028               1317 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male small intestine cDNA, RIKEN full-length
            enriched library, clone:2010002A20:Immunoglobulin domain containing
            protein, full insert sequence.
ACCESSION   AK008028
VERSION     AK008028.1  GI:12841960
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to
            mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2010002A20.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia, Eutheria, Rodentia, Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1317)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
```

```
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full length by cap trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 20.0. Second strand cDNA was prepared with the primer adapter
            of sequence[5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1317
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1894220"
                     /db_xref="taxon:10090"
                     /clone="2010002A20"
                     /sex="male"
                     /tissue_type="small intestine"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             85..846
                     /note="Immunoglobulin domain containing protein
                     data source:Pfam, source key:PF00047, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB25419.1"
                     /db_xref="GI:12841961"
                     /translation="MVWKITGPLQAVQLLEVVISLPQGRISSVILVNGRTLNYTLDLQ
                     HGVQASLECAVQNHTEDEELLWYREDGIVDLKNGNKINISSVCVSPINESDNGVRFTC
                     KLQRDQTVSVTVVLNVTFPPLLSGNGFQTVEENSDVSLVCNVKSNPQAQMMWYKNNSA
                     LVLEKGRHQTPQTRFSFQLSTTKVKKSDNGIYSCIASSSLKMEIMDFHLLVKDEVEVM
                     PAEPIIAACVVVVLTMAFALFSRRKRIMKFLWEKE"
     polyA_signal    1300..1305
                     /note="putative"
     polyA_site      1317
                     /note="putative"
BASE COUNT      401 a    252 c    304 g    360 t
ORIGIN
```

```
  Query Match             64.7%;  Score 487.2;  DB 11;  Length 1317;
  Best Local Similarity   78.9%;  Pred. No. 7.1e-129;
  Matches  590;  Conservative    2;  Mismatches  155;  Indels    1;  Gaps    1;

Qy      19 ctctctctaqqtaattttaatttctqcaccqtqaqatqacaaqttctqttttaaactqtqaat 78
           ||||||||||  |||||| | ||||||  | ||  ||||| ||||||||||||||||||||
Db     124 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGAGAAGAAGCTCTGTTTTAACTGTGAAT 183

Qy      79 qqtaaaactqaqaactatatctqqatactacacctqqctctccaaqcatctctqatatqt 138
           |||| ||||||||||||||| ||||| |||  || |||   ||||||||||||  ||||
Db     184 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 243

Qy     139 qctqttcaaaaccacacccaqaqaqqaaqaactqctctqqtaccqaqaqqqaqqqtttaqaqtq 198
           ||||||||||||||||||  || ||||||| ||||||||| |||| || || | |||
Db     244 GCTGTTCAAAACCACACCCAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 303

Qy     199 qatttqaaaatctqsaaacaaMatcaattccaqctctqtctqtctttcatcaqtqaa 258
           ||| ||||| ||:|||||:|||| ||| ||||| |||||||| ||||||  |||
Db     304 CATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 363

Qy     259 aatqacaacqqaatcaqcttttaactqcaqqctqqqqaqqqatcaqtccqtqtccqtttcq 318
           | |||||| ||| || |||||||||||| |||   | | |||||| | |||||||| |
Db     364 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 423

Qy     319 qtqqtqctqaatqttactttttctctctctaaqtqqaaacqactttcaaacaqttqaq 378
           || |||||||| |||| |||||||||||| |||||||| || | |||||||| || |||
Db     424 CTAACGCTCAACATTACCTTTTCTCTCTTCTAAGTGCCAATCGCTTCCAAACCGTCGAC 483

Qy     379 qaaqqcaqtaatqtqaaqttqqtttqcaatqtqaaaqccaaccccccaqqctcaaatqatq 438
           |||  |||  |||| |  |||||||||| |||||| ||||||| || || |||||||||
Db     484 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 543

Qy     439 tqqtacaaaaacaqtaqtctctctqatttaqaqaaaaqccqtcaccaaatccccccaqaca 498
           ||||| |||||||  ||   | | | |||||||||| ||||||||||||||  ||||||
Db     544 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCCCCCAGACA 603

Qy     499 aqtqaqtctttqcaqctqtcaatcaccaaaqtcaqaaaqtctqaccaatqqqacctacaqt 558
           || |||||||||||| |||| |||||||||||| ||||  |||||| || ||||||||
Db     604 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 663

Qy     559 tqtattqcaaaqtcatctctqaaaacqqaqaqcttqqactttcaccttqattqttaaagat 618
           |||||||| |  |||||| |||||| ||||| | ||||||| |||||| ||||||||||
Db     664 TCTATTGCCAGTCATCTTTTCAAAATCCACAGCATGCACTTCCACGTGCTTGTTAAAGAC 723

Qy     619 aaaactqtqqqtqtacaaataqaqcccattattqctqcatqtqttgtqatctttctqaca 678
           |||  | | |   | ||   |||||||| |||||||| ||||| ||| || | |||||
Db     724 AAAGTTTTTGTCATGCCGGCAGAGCCCATCATTGCTGCTTGTGTCGTGGTCGTCGTGACG 783

Qy     679 ttqtqctttqqactqattgctaqaaqaaaqaaaataatqaaq ctctqcatqaaqqqataa 737
           ||   |||| ||| || |||| || || | ||||||||||  | ||   ||| | ||
Db     784 ATGGCTTTTGCCCTGTTTTCTAGGAGGAAAAGAATAATGAAGTTTTTGTGGGAAAAAGAA 843

Qy     738 aqaccctcacaqtqaaacaqctctatqa 765
           |||||  |   ||||||||||||||| |
Db     844 TGACCCGAATTCAGAAACAGCTCTATAA 871
```

```
RESULT   4
BG970132
LOCUS       BG970132                 797 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602834263F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4973751
            5', mRNA sequence.
ACCESSION   BG970132
VERSION     BG970132.1  GI:14357769
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10962  row: m  column: 16
            High quality sequence stop: 795.
FEATURES             Location/Qualifiers
```

```
     source          1. .797
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4973751"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb  Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      249 a    174 c    181 g    193 t
ORIGIN


  Query Match             56.9%;  Score 435.4;  DB 10;  Length 797;
  Best Local Similarity   79.2%;  Pred. No. 4.8e-114;
  Matches  552;  Conservative    2;  Mismatches  138;  Indels    5;  Gaps    3;

Qy      33 tttatttctgccacgtgagatgacaagttctgttttaactgtgaatggtaaaaactgagaa 92
           ||||| | ||||||  | ||  ||||| ||||||||||||||||||||||| |||||||||
Db       1 TTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAATGGTAGAACTGAGAA 60

Qy      93 ctatatcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaacca 152
           |||||| ||||| |||  || |||   ||||||||||||||  ||||||||||||||||||
Db      61 CTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGTGCTGTTCAAAACCA 120

Qy     153 caccagagaggaagaactgctctggtaccgagaggagggagagtggatttgaaatctgs 212
           ||||   || |||||||| |||||||| |||| || || | ||| ||| |||||  ||:
Db     121 CACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTAGATCTGAAAAATGG 180

Qy     213 aaacaaatatcaattccagctctgtctgtgtctcttccatcagtgaaaatgacaacggaat 272
           ||||||:|||||   ||| |||| |||||||| |||||| |||| |||||| ||| |
Db     181 AAACAAAATCAACATCAGTTCTGTATGTGTCTCTCCCATCAACGAAAGTGACAATGGAGT 240

Qy     273 cagctttacctgcaggctggggagggatcagtccgtgtccgtttcggtggtgctgaatgt 332
           | |||||||||||| |||   | | |||||| | |||||||| | || ||||||||| ||
Db     241 CCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACAGTAGTGCTGAACGT 300

Qy     333 tacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggcagtaatgt 392
           ||| |||||||||| |||||||| || | |||||||| || |||||  |||  ||||
Db     301 TACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACAGCGATGT 360

Qy     393 gaagttggtttgcaatgtgaaagccaacccccaggctcaaatgatgtggtacaaaaaacag 452
           |  |||||||||| |||||| ||||||| || || ||||||||||||| |||||||
Db     361 AAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGGCAAATGATGTGGTATAAAAACAA 420

Qy     453 tagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtcttttca 512
           ||   | | | |||||||||| |||||||||||||| |||||||| ||||||||||
Db     421 CAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACAAGGGAGTCTTTTCA 480

Qy     513 gctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaagtc 572
           | |||| |||||||||||| |||| |||||| || |||||||| |||||||| |  ||
Db     481 GTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGCTGTATTGCCAGCTC 540

Qy     573 atctctgaaaacggagagcttggactttcacctgattgttaaagataaaactgtgggtgt 632
           |||||||||| ||||| | |||||| |||||| |||||||||  ||| | | |   |
Db     541 ATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGA-CAAAGTTTTGTCAT 599

Qy     633 accaatagagcccattattgctgcatgtgttgtga-tctttctgacattgtgctttggac 691
           ||   |||||||| |||||||| |||| |||| || | ||||     ||||| | |
Db     600 GCCGGCAGAGCCCATCATTGCTGCTTGTGTCGTGAGTCGTCCTGACGAATGGCTTTTGCC 659

Qy     692 ---tgattgctagaagaaagaaaatgaagctctg 725
               |  ||    || |||  | | ||||||||||||||
Db     660 CTGTTTTTCTAGGAGGAACAAGACTAATGAAGCTCTG 696


RESULT   5
AA986919
LOCUS       AA986919                 646 bp    mRNA    linear   EST 28-MAY-1998
DEFINITION  uc80b02.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:1431915 5' similar to TR:Q01974 Q01974 PROTEIN-TYROSINE
            KINASE TRANSMEMBRANE RECEPTOR ROR2 PRECURSOR ;, mRNA sequence.
ACCESSION   AA986919
VERSION     AA986919.1  GI:3168467
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:915983
            Possible reversed clone: similarity on wrong strand
            Seq primer: primer name ambiguous
            High quality sequence stop: 516.
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1431915"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG)  XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      192 a    145 c    146 g    163 t
ORIGIN


  Query Match             55.6%;  Score 425.6;  DB 9;  Length 646;
  Best Local Similarity   81.9%;  Pred. No. 3e-111;
  Matches  488;  Conservative    2;  Mismatches  106;  Indels    0;  Gaps    0;

Qy      19 cttctcttagtaatttttatttctgccacgtgaagatgacaagttctgttttaactgtgaat 78
           ||||||||||  |||||| | |||||  | ||  ||||| ||||||||||||||||||||
Db      50 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 109

Qy      79 ggtaaaaactgaagaactatatcctggatactacacctggctcccaagcatctctgatatgt 138
           |||| |||||||||||||| ||||| ||| || |||   ||||||||||||||  ||||
Db     110 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 169

Qy     139 gctgttcaaaaccacaccagagaggaagaaactgctctggtaccgagaggaggggagagtg 198
           |||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db     170 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 229

Qy     199 gatttgaaaatctgsaaacaaatcaattccagctctgtctgtgtctcttccatcagtgaa 258
```

```
Db    230 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 289

Qy    259 aatgacaacggaatcagctttacctgcaggctgggqaqggatcagtccgtgtccgtttcg 318
          | |||||| ||| || |||||||||||| |||   | | |||||| | |||||||| |
Db    290 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 349

Qy    319 gtggtgctgaatgttacttttcctcctctcctaaqtggaaacgacttccaaacagttgaq 378
          || |||||||| ||||| |||||||||||| |||||||| || | |||||||| || |||
Db    350 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 409

Qy    379 qaaqqcagtaatgtgaaqttqqtttqcaatgtgaaaqccaaccccccaggctcaaatgatg 438
          |||  |||  |||| |  |||||||||||| |||||| |||||| || || |||||||||
Db    410 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 469

Qy    439 tggtacaaaaacaqtaqtctcctcgatttagagaaaaqccgtcaccaaatccaacagaca 498
          ||||| ||||||| ||   | | | |||||||||| ||||||||||||||||| ||||||
Db    470 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACA 529

Qy    499 agtgaqtcttttcaqctgtcaatcaccaaaqtcqaqaaqcctgacaacgqaacctacaqt 558
          || |||||||||||| |||| |||||||||||| ||||  ||||||| || ||||||||
Db    530 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 589

Qy    559 tgtattgcaaaqtcatctctgaaaacggaqaqcttggactttcacctgattgttaa 614
          ||||||| |  |||||||||||| ||||| | ||||||| |||||| |||||||
Db    590 TGTATTGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAA 645
```

RESULT 6
BF785067
LOCUS BF785067 964 bp mRNA linear EST 12-JAN-2001
DEFINITION 602108226F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4236606 5', mRNA sequence.
ACCESSION BF785067
VERSION BF785067.1 GI:12090090
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 964)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9845 row: c column: 07
High quality sequence stop: 761.

FEATURES Location/Qualifiers
source 1..964
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4236606"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT 279 a 201 c 246 g 237 t 1 others
ORIGIN

Query Match 54.4%; Score 415.8; DB 10; Length 964;
Best Local Similarity 78.2%; Pred. No. 2.1e-108;

Matches 559; Conservative 2; Mismatches 149; Indels 5; Gaps 5;

```
Qy     19 cttctcttagtaatttttatttctgccacgtgagatgacaaqttctqttttaactqtqaat 78
          ||||||||||| |||||| | |||||| | || ||||| ||||||||||||||||||||
Db     35 CTTCTCTTAGTGGTTTTATCT-TGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 93

Qy     79 qgtaaaactgagaactatatcctggatactacacctggctcccaagcatctctgatatgt 138
          |||| |||||||||||||||| ||||| ||| || ||| ||||||||||||| ||||
Db     94 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 153

Qy    139 gctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggggagagtg 198
          |||||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db    154 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 213

Qy    199 gatttgaaatctgsaaacaamatcaattccaqctctgtctgtctcttccatcagtgaa 258
          ||| ||||| ||:|||||:|||||  ||| |||| |||||||||| ||||||  |||
Db    214 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGT-TGTGTCTCTCCCATCAACGAA 272

Qy    259 aatgacaacggaatcagctttacctgcaggctgggqaqggatcagtccgtgtccgtttcg 318
          | |||||| ||| || |||||||||||| |||   | | |||||| | |||||||| |
Db    273 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 332

Qy    319 gtggtgctgaatgttacttttcctcctctcctaaqtggaaacgacttccaaacagttgaq 378
          || |||||||| ||||| |||||||||||| |||||||| || | |||||||| || |||
Db    333 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 392

Qy    379 qaaqqcagtaatgtgaaqttqgtttqcaatgtgaaaqccaaccccccaggctcaaatgatg 438
          |||  |||  |||| |||||||||||| |||||| |||||| || || |||||||||
Db    393 GAAAACAGCGATGT-AAGTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 451

Qy    439 tggtacaaaaacaqtaqtctcctcgatttagagaaaaqccgtcaccaaatccaacagaca 498
          ||||| ||||||| ||   | | | |||||||||| ||||||||||||||||| ||||||
Db    452 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACA 511

Qy    499 agtgaqtcttttcaqctgtcaatcacc-aaaqtcgaqaaqcctgacaacgqaacctacaq 557
          || |||||||||||| |||| ||||| |||||| |||| ||||||| || ||||||||
Db    512 AGGGAGTCTTTTCAGTTGTCCATCACCAAAAGTCAAGAAATCTGACAATGGGACCTACAG 571

Qy    558 ttgtattgcaaaqtcatctctgaaaacggaqaqcttggactttcacctgattgttaaaqa 617
          |||||||  |  |||||||||||||| ||||| | ||||||| |||||| ||||||||||
Db    572 CTGTATTGGCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGA 631

Qy    618 taaaactgtqqgtqtaccaatagaqcccattattgctgcatgtgttgtgatctttctgac 677
          ||  | | |   | ||   |||||||| |||||||| ||||| |||     ||
Db    632 CCAACTTTTTCTCATCCCCCCAGAGCCCATCATTGCTGCTTGTGTCGTGGGTCGTC-CTC 690

Qy    678 attgtgctttggactgattgctagaaqaaaqaaaatatgaaqgctctgcatgaaq 732
          |   || ||||| ||| || |||| || || | |  |   |  || ||    |||
Db    691 ACGATGGTTTGGCCTGTTTTCTAGCAGGAAAAGATAATGAAGTCTGTGGGAAAAG 745
```

RESULT 7
BM445724
LOCUS BM445724 538 bp mRNA linear EST 05-FEB-2002
DEFINITION 1IL15A10.ab1 Bos taurus Ileum #1 library Bos taurus cDNA, mRNA sequence.
ACCESSION BM445724
VERSION BM445724.1 GI:18529880
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 538)
AUTHORS Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon,P.M.K. and Moore,S.S.
TITLE Gene Expression Profiling of the Bovine Gastrointestinal Tract
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Stephen Moore
. Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For. Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada

```
            Tel: 780 492 0169
            Fax: 780 492 4265
            Email: smoore@afns.ualberta.ca
            Insert Length: 538   Std Error: 0.00
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .538
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="Bos taurus Ileum #1 library"
                      /tissue_type="Smooth muscle"
                      /cell_type="Simple columnar epithelial"
                      /dev_stage="Young adult"
                      /lab_host="XL1-BlueMRF'strain"
                      /note="Organ: Intestine/ileum; Vector: Uni-2ZAPXR; Site_1:
                      EcoRI; Site_2: Xho I"
BASE COUNT      165 a    118 c    121 g    134 t
ORIGIN


  Query Match             52.6%;  Score 402.6;  DB 10;  Length 538;
  Best Local Similarity   85.7%;  Pred. No. 1.1e-104;
  Matches  457;  Conservative    2;  Mismatches   71;  Indels    3;  Gaps    1;

Qy      42 gccacgtgagatgacaagttctgttttaactgtgaatggtaaaactgagaactatatcct 101
           ||||| || |||||||||||||||||||| |||||||||||||||||||| |||||||| ||
Db       5 GCCACATGTGATGACAAGTTCTGTTTTAAGTGTGAATGGTAAAACTGAAAACTATATTCT 64

Qy     102 ggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacaccagaga 161
           ||| |||   ||||||  ||||| |||||||| |||||||||||||||||||| | ||||
Db      65 GGACACTGAGCCTGGCCTCCAAGAATCTCTGAAATGTGCTGTTCAAAACCACATCCGAGA 124

Qy     162 ggaagaactgctctggtaccgagaggaggggagagtggatttgaaatctgsaaacaamat 221
           |||||||| |||||||||| || || |||||||||||||||| ||||:|||||||:||
Db     125 TGAAGAACTTCTCTGGTACCGCGAAGATGGGAGAGTGGATTTGAAGTCTGGAAACAAAAT 184

Qy     222 caattccagctctgtctgtgtctcttccatcagtgaaaatgacaacggaatcagctttac 281
           ||| ||||||||||||||||||   |||||||||| |||||||||| |||| ||||||
Db     185 CAACTCCAGCTCTGTCTGTGTCTCGGGCATCAGTGAAGATGACAACGGCATCACCTTTAC 244

Qy     282 ctgcaggctggggagggatcagtccgtgtccgtttcggtggtgctgaatgttacttttcc 341
           ||||| |||| |||| ||||||| |||||| | || |||||||||||| || ||||||||
Db     245 CTGCAAGCTGCAGAGGAATCAGTCGGTGTCCATCTCAGTGGTGCTGAACGTCACTTTTCC 304

Qy     342 tcctctcctaagtggaaacgacttccaaacagttgaggaaggcagtaatgtgaagttggt 401
           |||||||||||||||||| |||||||||||| |||||||||||| ||||||||||||
Db     305 TCCTCTCCTAAGTGGAAATGACTTCCAAACAGCTGAGGAAGGCAGTGATGTGAAGTTGGT 364

Qy     402 ttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtacaaaaacagtagtctcct 461
           |||||||||||| |||||||||||||||| |||||||||| |||||| | |  || |
Db     365 TTGCAATGTGAAATCCAACCCCCAGGCTCAGATGATGTGGTATAAAAACAATGGCATCTT 424

Qy     462 cgatttagagaaaagccgtcaccaaatccaacagacaagtgagtcttttcagctgtcaat 521
           | ||||||||   || ||||||||||||||||||||| ||| |||||  |||| ||
Db     425 GAACTTAGAGAA---CCATCACCAAATCCAACAGACAAGCAAGTAGTTTCAATTGTCGAT 481

Qy     522 caccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaagtcat 574
           |||||||||| ||||  | ||||| || || || || |||||||| || ||||
Db     482 CACCAAAGTCAAGAAATCCGACAATGGGACTTATAGCTGTATTGCCAATTCAT 534


RESULT   8
BF788117
LOCUS       BF788117                936 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  602113341F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4241194
            5', mRNA sequence.
ACCESSION   BF788117
VERSION     BF788117.1  GI:12093166
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 936)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement. Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution. MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9857  row: b  column: 11
            High quality sequence start: 5
            High quality sequence stop: 713.
FEATURES              Location/Qualifiers
     source           1. .936
                      /organism="Mus musculus"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:4241194"
                      /clone_lib="NCI_CGAP_Kid14"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.75 kb. Constructed by Life
                      Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      270 a    213 c    231 g    220 t      2 others
ORIGIN


  Query Match             50.0%;  Score 382.2;  DB 10;  Length 936;
  Best Local Similarity   80.8%;  Pred. No. 9.4e-99;
  Matches  492;  Conservative    2;  Mismatches  110;  Indels    5;  Gaps    4;

Qy      19 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 78
           ||||||||||  |||||| | ||||||  | ||  ||||| || ||||||||||||||||
Db      31 CTTCTCTTAGTGGTTTTATCT-TGCCACAGGGGAGAACAAGCTC-GTTTTAACTGTGAAT 88

Qy      79 ggtaaaactgagaactatatcctggatactacacctggctcccaagcatctctgatatgt 138
           |||| ||||||||||||||| ||||| |||  || |||   ||||||||||||  ||||
Db      89 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 148

Qy     139 gctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggggagagtg 198
           ||||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db     149 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 208

Qy     199 gatttgaaatctgsaaacaamatcaattccagctctgtctgtgtctcttccatcagtgaa 258
           ||| |||||  ||:||||:||||    ||| ||||| |||||||| ||||||   |||
Db     209 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGT-TGTGTCTCTCCCATCAACGAA 267

Qy     259 aatgacaacggaatcagctttacctgcaggctggggagggatcagtccgtgtccgtttcg 318
           | |||||| ||| || |||||||||||| |||   | | |||||| | |||||||| |
Db     268 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 327

Qy     319 gtggtgctgaatgttacttttcctcctcctaagtggaaacgacttccaaacagttgag 378
           || |||||||| |||||| |||||||||||| |||||||| || | |||||||| || |||
Db     328 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 387

Qy     379 gaaggcagtaatgtgaagttggtttgcaatgtgaaaagccaaccccccaggctcaaatgatg 438
           |||  |||  |||| |  | | |||||| |||||| |||||| || || |||||||||
Db     388 GAAAACAGCGATGT--AAGTTGGTTGCAAGGTGAAATCCAACCCTCAAGCGCAAATGATG 445

Qy     439 tggtacaaaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagaca 498
           ||||| |||||||  ||   | | | |||||||||| ||||||||||||||||| ||||||
Db     446 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGCCGTCACCAAATCCACCAGACA 505

Qy     499 agtgagtcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagt 558
           || |||||||||| |||| ||||||||||| |||  ||||||| || ||||||||
Db     506 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 565

Qy     559 tgtattgcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagat 618
```

```
          ||||| || |  ||||||||||||| ||||| | ||||||| |||||| ||||||||||
Db      566 TGTATCGCCAGCTCATCTCTGAAAATGGACACCATGCACTTCCACCTGCTTGTTAAAGAC 625

Qy      619 aaaactgtg 627
            ||| | ||
Db      626 AAAGTTTTG 634


RESULT   9
AA500290
LOCUS       AA500290                 543 bp    mRNA    linear   EST 01-JUL-1997
DEFINITION  vi97g08.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:920222 5' similar to TR:G337467 G337467
            PROTEIN TYROSINE KINASE TRANSMEMBRANE RECEPTOR ROR2 PRECURSOR ,,
            mRNA sequence.
ACCESSION   AA500290
VERSION     AA500290.1  GI:2235257
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,P.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL , contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:582438
            Possible reversed clone, similarity on wrong strand
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 431.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Mus musculus"
                     /strain "FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:920222"
                     /clone_lib="Barstead mouse pooled organs MPLRB4"
                     /sex="mixed"
                     /tissue_type "pooled organs"
                     /dev_stage="7 day"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker, Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [GTTGGATTCGGTACC], digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead."
BASE COUNT      164 a    123 c    121 g    135 t
ORIGIN


  Query Match             49.9%;  Score 381.8;  DB 9;  Length 543;
  Best Local Similarity   81.3%;  Pred. No. 1.1e-98;
  Matches  440;  Conservative    2;  Mismatches   99;  Indels    0;  Gaps    0;

Qy      41 tgccacgtgagatgacaagttctgttttaactgtgaatggtaaaactgagaactatatcc 100
           |||||| || ||  ||||| ||||||||||||||||||||| ||||||||||||||| |
Db       2 TGCCACATGGGAGAACAAGCTCTGTTTTAACTGTGAATGGTAGAACTGAGAACTATATTC 61

Qy     101 tggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacaccagag 160
           |||| ||| || |||  ||||||||||||| ||||||||||||||||||||||| |
Db      62 TGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGTGCTGTTCAAAACCACACCAGG 121

Qy     161 agggaagaactgctctggtaccgagaggagggggagagtggatttgaaatctgsaaacaama 220
           | ||||| || |||||||| |||| || || | ||| ||| |||||  ||:|||||:|
Db     122 ATGAAGATCTTCTCTGGTACAGACAAGATGGAATAGTAGATCTGAAAAATGGAAACAAAA 181

Qy     221 tcaattccagctctgtctgtgtcttccatcagtgaaaatgacaacggaatcagcttta 280
           ||||   ||| ||||| |||||||| |||||  |||| |||||| ||| || ||||||
Db     182 TCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAAAGTGACAATGGAGTCCGCTTTA 241

Qy     281 cctgcaggctgggggagggatcagtccgtgtccgtttcgggtggtgctgaatgttacttttc 340
           |||||| |||   | | |||||| | ||||||||| | || |||||||| ||||| ||||
Db     242 CCTGCAAGCTACAGCCGGATCAGACGGTGTCCGTTACAGTAGTGCTGAACGTTACTTTC 301

Qy     341 ctctctctctaagtggaaacgacttccaaacagttgaggaaggcagtaatatgtgaagttgg 400
           ||||||| |||||||| || | |||||||| || |||||  |||  |||| | ||||
Db     302 CTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACAGCGATGTAAGTTTGG 361

Qy     401 tttgcaatgtgaaagcaaccccaggctcaaatgatgtggtacaaaaacagtagtctcc 460
           ||||||| |||||| ||||||| || || |||||||||||||| |||||||  ||   |
Db     362 TTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATGTGGTATAAAAACAACAGCGCCT 421

Qy     461 tcgatttgggggaaaagccgtcaccaaatccaacagacaaagtgagtcttttcag.tgtcaa 520
           | | |||||||||| |||||||||| ||||| |||||||| |||||||||||| |||| |
Db     422 TGGTTTTAGAGAAAGGCCGTCACCACATCCACCAGACAAGGGAGTCTTTTCAGTTGTCCA 481

Qy     521 tcaccaaagtcgagaaagcctgacaacgqaacctacagttgtattgcaaaqtcatctctga 580
           |||||||||| |||| |||||| || ||||||| ||||||| |  |||||||||
Db     482 TCACCAAAGTCAAGAAATCTGACATTGGGACCTACAGCTGTATTGCCAGCTCATCTCTGA 541

Qy     581 a 581
           |
Db     542 A 542


RESULT  10
A1747948
LOCUS       A1747948                 657 bp    mRNA    linear   EST 22-JUN-1999
DEFINITION  ul04c10.y1 Soares mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2065074 5' similar to TR:Q01974 Q01974 PROTEIN-TYROSINE
            KINASE TRANSMEMBRANE RECEPTOR ROR2 PRECURSOR ,, mRNA sequence.
ACCESSION   A1747948
VERSION     A1747948.1  GI:5126212
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Rock,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Papé,D., Harvey,N., Schark,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:994262
            Seq primer: custom primer used
            High quality sequence stop: 520.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Mus musculus"
```

```
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2065074"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      195 a    151 c    153 g    158 t
ORIGIN


  Query Match           43.9%;  Score 336.2;  DB 9;  Length 657;
  Best Local Similarity   78.8%;  Pred. No. 1.4e-85;
  Matches  401;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy     257 aaaatgacaacggaatcagctttacctgcaggctggggagggatcagtccgtgtccgttt 316
           | |||||||| ||| || |||||||||||| |||   | | |||||| | |||||||||
Db      82 ACAATGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTA 141

Qy     317 cggtggtgctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttg 376
           | || |||||||| ||||| |||||||||||| |||||||| || | |||||||||| || |
Db     142 CAGTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGG 201

Qy     377 aggaaggcagtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatga 436
           |||||  |||  |||| |  |||||||||| |||||| ||||||| || || |||||||
Db     202 AGGAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGA 261

Qy     437 tgtggtacaaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacaga 496
           ||||||| |||||||  ||   | | | |||||||||| ||||||||||||||| ||||
Db     262 TGTGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGA 321

Qy     497 caagtgagtcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctaca 556
           |||| |||||||||||| |||| |||||||||||| ||||  ||||||| || |||||||
Db     322 CAAGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACA 381

Qy     557 gttgtattgcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaag 616
           | ||||||||  |  |||||||||||| ||||| | ||||||| |||||| |||||||||
Db     382 GCTGTATTGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAG 441

Qy     617 ataaaactgtgggtgtaccaatagagcccattattgctgcatgtgttgtgatctttctga 676
           | |||  | | |   | ||  |||||||||| |||||||| ||||| ||| || | ||||
Db     442 ACAAAGTTTTTGTCATGCCCGCAGAGCCCATCATTGCTGCTTGTGTCGTGGTCGTCCTGA 501

Qy     677 cattgtgctttggactgattgctagaagaaagaaaataatgaagctctgcatgaaggata 736
           |  ||   ||||  |||| || |||| || || | ||||||||||||||   |||  | |
Db     502 CGATGGCTTTTGCCCTGTTTTCTAGGAGGAAAAGAATAATGAAGCTCTGTGGGAAAAAGA 561

Qy     737 aagaccctcacagtgaaacagctctatga 765
           | |||||  |   ||||||||||||| |
Db     562 ATGACCCGAATTCAGAAACAGCTCTATAA 590


RESULT  11
AI956638
LOCUS       AI956638                445 bp    mRNA    linear   EST 20-AUG-1999
DEFINITION  ul79f12.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2136815 5' similar to SW:VGR2_HUMAN P35968 VASCULAR
            ENDOTHELIAL GROWTH FACTOR RECEPTOR 2 PRECURSOR ;, mRNA sequence
ACCESSION   AI956638
VERSION     AI956638.1  GI:5749347
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,R., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1001491
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2136815"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      136 a     94 c    102 g    113 t
ORIGIN


  Query Match           40.1%;  Score 306.6;  DB 9;  Length 445;
  Best Local Similarity   80.2%;  Pred. No. 4.1e-77;
  Matches  357;  Conservative    2;  Mismatches   86;  Indels    0;  Gaps    0;

Qy      33 tttatttctgcaagctgaggatgacaagttctgttttaactgtgaatggtaaaactgagaa 92
           ||||| | |||||  | ||  ||||| ||||||||||||||||||||||| |||||||||
Db       1 TTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAATGGTAGAACTGAGAA 60

Qy      93 ctatatcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaacca 152
           |||||| ||||| |||  || |||   |||||||||||||  ||||||||||||||||||
Db      61 CTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGTGCTGTTCAAAACCA 120

Qy     153 caccagagaggaaagaactgctctggtaccgagaagaaggggagaagtggatttgaaatctgs 212
           ||||   || |||||||| ||||||||| |||| || || | ||| ||| |||||  ||:
Db     121 CACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTAGATCTGAAAAATGG 180

Qy     213 aaacaamatcaattccagctctgtctgtgtctcttccatcagtgaaaatgacaacggaat 272
           ||||||:|||||   ||| ||||| |||||||| ||||||  |||| |||||| ||| |
Db     181 AAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAAAGTGACAATGGAGT 240

Qy     273 cagctttacctgcaggctggggagggatcagtccgtgtccgtttcggtggtgctgaatgt 332
```

```
          | ||||||||||||| |||   | | |||||| | ||||||||| | || |||||||| ||
Db    241 CCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACAGTAGTGCTGAACGT 300

Qy    333 tacttttcctcctctcctaaqtqaaacqacttccaaacaqttqaggaaqqcaqtaatqt 392
          ||| |||||||||| ||||||| || | ||||||||| || |||||  |||  |||| 
Db    301 TACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACAGCGATGT 360

Qy    393 qaaqttqqtttqcaatqtqaaaqccaaccccccaqqctcaaatgatgtggtacaaaaacaq 452
          |  |||||||||||  |||||| ||||||| || || ||||||||||||||| |||||||
Db    361 AAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATGTGGTATAAAAACAA 420

Qy    453 taqtctcctcqatttagaqaaaaqc 477
            ||   | | | |||||||||| ||
Db    421 CAGCGCCTTGGTTTTAGACAAAGGC 445


RESULT  12
AI615364
LOCUS       AI615364                487 bp    mRNA    linear   EST 21-APR-1999
DEFINITION  vi97q08.y1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:920222 5', mRNA sequence.
ACCESSION   AI615364
VERSION     AI615364.1  GI:4624581
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty free through LLNL , contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:532438
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 478.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:920222"
                     /clone_lib="Barstead mouse pooled organs MPLRB4"
                     /sex="mixed"
                     /tissue_type="pooled organs"
                     /dev_stage="7 day"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [GTTGGATTCGGTACC], digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead."
BASE COUNT      146 a    109 c    111 g    120 t      1 others
ORIGIN


  Query Match           38.6%,  Score 295.6;  DB 9;  Length 487;
  Best Local Similarity   80.8%;  Pred. No. 6e-74;
  Matches  354;  Conservative    2;  Mismatches   81;  Indels    1;  Gaps    1;

Qy      19 cttctcttagtaatttttatttctgccacgtqaqatqacaaqttctqttttaactqtqaat 78
           ||||||||||  |||||| | ||||||  | ||  ||||| ||||||||||||||||||
Db      47 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGAGAACAAGCTCTGTTTTAACTGTGAAT 106

Qy      79 qqtaaaactqaqaactatatcctqqatactacacctqqctcccaaqcatctctqatatqt 138
           |||| |||||||||||||||| ||||| |||  || |||   |||||||||||||  |||| 
Db     107 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 166

Qy     139 qctqttcaaaaccacaccaqaqaqqaaqaactqctctqqtaccqaqaqqaqqqqaqaqtq 198
           |||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db     167 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 226

Qy     199 gatttgaaaatctqsaaaacaamatcaattccaqctctqtctqtctcttccatcaqtgaa 258
           ||| |||||  ||:||||||:|||||   ||| ||||| |||||||| |||||  |||
Db     227 CATCTCAAAAATGGAAACAAAATCAACATCACTTCTGTTTGTGTCTCTCCCATCAACGAA 286

Qy     259 aatqaqcaacqqaatcaqctttacctqcaqqctqqqqaqqqatcaqtccqtqtccqtttcq 318
           | |||||| ||| || ||||||||||||| |||   | | |||||| | |||||||| |
Db     287 ACTCACAATCCAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 346

Qy     319 qtqqtqctqaatqttacttttcctcctctcctaaqtqqaaacqacttccaaacaqttqaq 378
           || |||||||| |||| |||||||||| ||||||| || | |||||||| || |||
Db     347 CTACTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 406

Qy     379 qaaqqcaqtaatqtqaaqttqqtttqcaatqtqaaaqccaaccccccaqqctcaaatqa-t 437
           |||  |||  |||| |  |||||||||| |||||| ||||||| || || ||||||| |
Db     407 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGANT 466

Qy     438 qtqqtacaaaaacaqtaq 455
           |||||| |||||||  ||
Db     467 GTGGTATAAAAACAACAG 484


RESULT  13
AW211197
LOCUS       AW211197                514 bp    mRNA    linear   EST 05-DEC-1999
DEFINITION  ul25h03.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2099381 5' similar to TR:060469 060469 DOWN SYNDROME CELL
            ADHESION MOLECULE. ,, mRNA sequence.
ACCESSION   AW211197
VERSION     AW211197.1  GI:6517137
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_EsIs: ul25h03.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:996313
            Seq primer: custom primer used
            High quality sequence stop: 513.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Mus musculus"
```

/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2099381"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTCCG and 3' end primer CGACCTGCAGCTCGAGCACA."
BASE COUNT 150 a 112 c 123 g 129 t
ORIGIN

Query Match 31.5%; Score 241.2; DB 9; Length 514;
Best Local Similarity 78.3%; Pred. No. 2.6e-58;
Matches 299; Conservative 2; Mismatches 80; Indels 1; Gaps 1;

```
Qy      19 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 78
           ||||||||||| |||||| | |||||| | ||  ||||| ||||||||||||||||| |
Db     133 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAGT 192

Qy      79 ggtaaaactgagaactatatcctggatactacacctggctcccaagcatctctgatatgt 138
           |||| ||||||||||||||| ||||| ||| || |||   |||||||||||| ||||
Db     193 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 252

Qy     139 gctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggggagagtg 198
           ||||||||||||||||   || |||||||| ||||||||| |||| || || | |||
Db     253 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 312

Qy     199 gatttgaaatctgsaaacaamatcaattccagctctgtctgtgtctcttccatcagtgaa 258
           ||| ||||| ||:||||||:||||   ||| ||||| |||||||| |||||| |||
Db     313 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 372

Qy     259 aatgacaacggaatcagctttacctgcaggct-ggggagggatcagtccgtgtccgtttc 317
           | |||||| ||| || |||||||||||| |||    | | |||||| |||||||| | |
Db     373 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACCAGCGCGATCAGACCGTGTCCCGTAC 432

Qy     318 ggtggtgctgaatgttactttcctcctctcctaagtggaaacgacttccaaacagttga 377
           || |||||  |   ||| |||||||||| |||||||| || | ||| ||||| || ||
Db     433 AGTAGTGCTTGACCGTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTTCAAACCGTGGA 492

Qy     378 ggaaggcagtaatgtgaagttg 399
           ||||  |||  |||| |  |||
Db     493 GGAAAACAGCGATGTAAGTTTG 514
```

RESULT 14
BB845854
LOCUS BB845854 437 bp mRNA linear EST 26-NOV-2001
DEFINITION BB845854 RIKEN full-length enriched, adult male kidney Mus musculus cDNA clone F520008G20 5', mRNA sequence.
ACCESSION BB845854
VERSION BB845854.1 GI:17084229
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi, Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 437)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okagaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1..437
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="F520008G20"
/clone_lib="RIKEN full length enriched, adult male kidney"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. "
BASE COUNT 125 a 99 c 99 g 114 t
ORIGIN

Query Match 28.2%; Score 215.4; DB 9; Length 437;
Best Local Similarity 79.8%; Pred. No. 6.3e-51;
Matches 264; Conservative 2; Mismatches 63; Indels 2; Gaps 1;

```
Qy      19 cttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaactgtgaat 78
           ||||||||||| |||||| | |||||| | ||  ||||| ||||||||||||||||| |
Db     107 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 166
```

```
Qy      79 qqtaaaactqaqaactatatcctqqatactacacctqqctcccaaqcatctctgatatqt 138
           |||| ||||||||||||||| ||||| ||| || |||   ||||||||||||  ||||
Db     167 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTCGAATCT 226

Qy     139 qctqttcaaaaccacacccaqaqaqqaaqaactqctctqqtaccqagaggagggqagaqtq 198
           |||||||||||||||||    || |||||||| |||||||| |||| || || | |||
Db     227 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 286

Qy     199 qatttqaaatctqsaaacaamatcaattccaqctctqtctqtqtctcttccatcaqtqaa 258
           ||| |||||  ||:|||||:|||||   ||| ||||| |||||||| ||||||  |||
Db     287 GATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 346

Qy     259 aatqacaacqqaatcaqc--tttacctqcaqqctqqqqaqqqatcaqtccqtqtccqttt 316
           | |||||| ||| || ||  ||  | |||| |||   | | ||||| | |||||||||
Db     347 AGTGACAATGGAGTCCGCCTTTACCGTGCAAGCTACAGCGCGATCAGACGGTGTCGGTTA 406

Qy     317 cqqtqqtqctqaatgttacttttcctcctct 347
           | || |||||||| |||| ||||||||||||
Db     407 CAGTAGTGCTGAACGTTACCTTTCCTCCTCT 437


RESULT 15
AZ864346/c
LOCUS       AZ864346                515 bp    DNA     linear   GSS 21-FEB-2001
DEFINITION  2M0173F20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0173F20 R, DNA sequence.
ACCESSION   AZ864346
VERSION     AZ864346.1  GI:14063555
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,R., Longacre,S., Mahmoud,M., Meenan,F., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000  Std Error: 0.00
            Plate: 0173  row: F  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 515.
FEATURES             Location/Qualifiers
     source          1..515
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0173F20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      138 a    117 c    105 g    155 t
ORIGIN


  Query Match             24.8%;  Score 189.8;  DB 12;  Length 515;
  Best Local Similarity   81.0%;  Pred. No. 1.5e-43;
  Matches  221,  Conservative    0,  Mismatches   52;  Indels    0;  Gaps    0;

Qy     354 tqqaaaacqaettccaaacaqttqaqqaaqqcaqtaatqtqaaqttqqtttqcaatqtqaa 413
           ||| || | |||||||||| || ||||||  |||  |||| |  ||||||||||| |||||
Db     515 TGGCAATGGCTTCCAAACCGTGGAGCAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAA 456

Qy     414 aqccaaccccaggctcaaatgatgtqqtacaaaaacagtagtctcctcgatttaqaqaa 473
           | ||||||| || || |||||||||||||| |||||||  ||     | | |||||||||
Db     455 AICCAACCCTCAAGCGCAAATGATGTGGTATAAAAACAACAGCGCTTTGGTTTTAGAGAA 396

Qy     474 aaqccqtcaccaaatccaacaqacaaqtqagtcttttcaqctqtcaatcaccaaaqtcga 533
           | |||||||||||||||| |||||||| |||||||||||| |||| ||||||||||| |
Db     395 AGGCCGTCACCAAATCCACCAGACAAGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAA 336

Qy     534 qaaqcctqacaacqqaacctacaqttgtattqcaaaqtcatctctqaaaacqqaqaqctt 593
           |||  |||||| || || |||| |||||||| |  ||||||||||| |||| | |
Db     335 GAAAICTGACAATGGGACTTACAGCTGTATTGCCAGCTCATCTCTGAAAATGGAGACCAT 276

Qy     594 qqactttcacctgattgttaaaqataaaactgt 626
           |||||| ||| || |||||||| ||| || |
Db     275 CCACTTCCACTTGCTTGTTAAAGGTAACTCTCT 243


Search completed: September 17, 2002, 09:04:18
Job time: 2571 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:09:58 ; Search time 102.65 Seconds
(without alignments)
428.063 Million cell updates/sec

Title: US-09-863-823-2
Perfect score. 1295
Sequence: 1 MQMGRFLLLLVILFLPREMTS.....RKKIMKLCMKDKDPHSETAL 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	966	74.6	261	11	Q9D7L8	Q9d718 mus musculu
2	915	70.7	253	11	Q9D8H2	Q9d8h2 mus musculu
3	175	13.5	604	4	Q96CJ3	Q96cj3 homo sapien
4	173.5	13.4	2762	5	P91255	P91255 caenorhabdi
5	172	13.3	2013	11	Q9ERC8	Q9erc8 mus musculu
6	171.5	13.2	946	13	Q07153	Q07153 torpedo cal
7	169.5	13.1	359	5	Q9V6C2	Q9v6c2 drosophila
8	169.5	13.1	431	4	Q9NYZ4	Q9nyz4 homo sapien
9	166	12.8	605	11	Q921P2	Q921p2 mus musculu
10	161	12.4	1011	5	Q24273	Q24273 drosophila
11	161	12.4	1079	5	Q9VNP2	Q9vnp2 drosophila
12	159	12.3	597	6	Q95LH0	Q95lh0 pan troglod
13	158	12.2	1463	11	O55124	O55124 mus musculu
14	157.5	12.2	1415	5	Q94155	Q94155 caenorhabdi
15	157	12.1	467	5	Q9VP08	Q9vp08 drosophila
16	156	12.0	1173	5	P91456	P91456 caenorhabdi
17	155	12.0	595	4	Q96PQ1	Q96pq1 homo sapien
18	154.5	11.9	1513	13	Q90Z70	Q90z70 brachydanio
19	154	11.9	509	11	Q9EQY5	Q9eqy5 mus musculu
20	154	11.9	509	11	Q91YK7	Q91yk7 mus musculu
21	154	11.9	603	5	Q9NKF5	Q9nkf5 drosophila
22	154	11.9	868	11	Q62838	Q62838 rattus norv
23	153.5	11.9	1344	11	Q9Z2I4	Q9z2i4 mus musculu
24	153	11.8	467	4	Q9Y286	Q9y286 homo sapien
25	152.5	11.8	850	4	Q9ULF7	Q9ulf7 homo sapien
26	152.5	11.8	5198	5	O76518	O76518 caenorhabdi
27	152	11.7	869	4	O15146	O15146 homo sapien
28	149	11.5	860	11	Q61095	Q61095 mus musculu
29	149	11.5	868	11	Q61006	Q61006 mus musculu
30	149	11.5	2037	5	Q9VIS8	Q9vis8 drosophila
31	148.5	11.5	4370	4	Q9H3V5	Q9h3v5 homo sapien
32	147.5	11.4	496	5	Q94224	Q94224 caenorhabdi
33	147	11.4	1445	11	Q63155	Q63155 rattus norv
34	146	11.3	528	5	P91670	P91670 drosophila
35	146	11.3	545	5	Q9VCT4	Q9vct4 drosophila
36	146	11.3	7962	4	Q10465	Q10465 homo sapien
37	145.5	11.2	773	4	Q9NSW7	Q9nsw7 homo sapien
38	145.5	11.2	1215	5	Q9V787	Q9v787 drosophila
39	145	11.2	951	5	Q19128	Q19128 caenorhabdi
40	144.5	11.2	334	7	O02870	O02870 gallus gall
41	144.5	11.2	350	7	O02869	O02869 gallus gall
42	144.5	11.2	693	4	Q9UPU1	Q9upu1 homo sapien
43	144.5	11.2	1482	5	Q9V4Y0	Q9v4y0 drosophila
44	144.5	11.2	6620	4	Q96AA2	Q96aa2 homo sapien
45	144	11.1	454	11	Q91W54	Q91w54 mus musculu

ALIGNMENTS

RESULT 1
Q9D7L8
ID Q9D7L8 PRELIMINARY; PRT; 261 AA.
AC Q9D7L8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 2010002A20RIK PROTEIN.
GN 2010002A20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK009118; BAB26083.1; -.
DR MGD; MGI:1913851; 2010002A20Rik.

```
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00409; IG; 2.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   261 AA;  29050 MW;  1FE1FC02D27E83CD CRC64;


  Query Match             74.6%;  Score 966;  DB 11;  Length 261;
  Best Local Similarity   75.8%;  Pred. No. 1.8e-76;
  Matches  188;  Conservative   25;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       7 LLLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRV 66
           ||||:| ||:  |||||||||:||||||||  | |||| |||||||| :|||||||||:| |
Db      14 LLLVVLSLPQGRTSSVLTVNGRTENYILDTQHGVQASLECAVQNHTEDEELLWYREDGIV 73

Qy      67 DLKSGNKINSSSVCVSSISENDNGISFTCRLGRDQSVSVSVVLNVTFPPLLSGNDFQTVE 126
           |||:|||||  |||||| |:|:|||: |||:| |||:|||:|||||||||||||||| |||||
Db      74 DLKNGNKINISSVCVSPINESDNGVRFTCKLQRDQTVSVTVVLNVTFPPLLSGNGFQTVE 133

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYS 186
           | |:| ||||||:||||||||||:| | ||| |||| || ||||||||||:| ||||||
Db     134 ENSDVSLVCNVKSNPQAQMMWYKNNSALVLEKGRHQIHQTRESFQLSITKVKKSDNGTYS 193

Qy     187 CIAKSSLKTESLDFHLIVKDKTVGVPIEPIIAACVVIFLTLCFGLIARRKKIMKLCMKDK 246
           ||| |||| |::|||||:||||   :| |||||||||: ||: | | :|||:|||| |
Db     194 CIASSSLKMETMDFHLLVKDKVFVMPAEPIIAACVVVVLTMAFALFSRRKRIMKLCGKKN 253

Qy     247 DPHSETAL 254
           ||:|||||
Db     254 DPNSETAL 261


RESULT   2
Q9D8H2
ID   Q9D8H2      PRELIMINARY;      PRT;   253 AA.
AC   Q9D8H2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   2010002A20RIK PROTEIN.
GN   2010002A20RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK008028; BAB25419.1; -.
DR   MGD; MGI:1913851; 2010002A20Rik.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00409; IG; 2.
DR   SMART; SM00408; IGc2; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   253 AA;  28372 MW;  BCA6EC8C8D7167D1 CRC64;


  Query Match             70.7%;  Score 915;  DB 11;  Length 253;
  Best Local Similarity   76.1%;  Pred. No. 5e-72;
  Matches  178;  Conservative   24;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       7 LLLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRV 66
           ||||:| ||:  |||||||||:||||||||  | |||| |||||||| :|||||||||:| |
Db      14 LLLVVLSLPQGRTSSVLTVNGRTENYILDTQHGVQASLECAVQNHTEDEELLWYREDGIV 73

Qy      67 DLKSGNKINSSSVCVSSISENDNGISFTCRLGRDQSVSVSVVLNVTFPPLLSGNDFQTVE 126
           |||:|||||  |||||| |:|:|||: |||:| |||:|||:|||||||||||||||| |||||
Db      74 DLKNGNKINISSVCVSPINESDNGVRFTCKLQRDQTVSVTVVLNVTFPPLLSGNGFQTVE 133

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYS 186
           | |:| ||||||:||||||||||:| | ||| |||| || ||||||||||:| ||||||
Db     134 ENSDVSLVCNVKSNPQAQMMWYKNNSALVLEKGRHQIHQTRESFQLSITKVKKSDNGTYS 193

Qy     187 CIAKSSLKTESLDFHLIVKDKTVGVPIEPIIAACVVIFLTLCFGLIARRKKIMK 240
           ||| |||| |::|||||:||||   :| |||||||||: ||: | | :|||:|||
Db     194 CIASSSLKMETMDFHLLVKDKVFVMPAEPIIAACVVVVLTMAFALFSRRKRIMK 247


RESULT   3
Q96CJ3
ID   Q96CJ3      PRELIMINARY;      PRT;   604 AA.
AC   Q96CJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO NEURAL CELL ADHESION MOLECULE 1 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY, AND RENAL CELL ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC014205; AAH14205.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   604 AA;  66271 MW;  B3B791A4D033D546 CRC64;


  Query Match             13.5%;  Score 175;  DB 4;  Length 604;
  Best Local Similarity   28.4%;  Pred. No. 5.3e-07;
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      41 QASLIC---------      AVQNHTREEELLWYREEGRVDLKSGNKINS    SSVCVS 82
           | :| |               : :| : ||:  | | | :  |     : |    ||: :
Db      70 QVTLICEASGDPIPSITWRTSTRNISSEEKASWTRPEKQETLDGHMVVRSHARVSSLTLK 129

Qy      83 SISENDNGISFIC----KLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 138
           ||   | |  : |     :|:|   | |: | | : | | |       ||: | : | |
Db     130 SIQYTDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 185

Qy     139 ANPQAQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTES 197
           | | | : |::  ||     |  :| | : | :|   : | | |:| | : :  ||
Db     186 AYPSAITSWFRDGQLLPSSNYSNIKIYNIPSASYLEVIPDSENDFGNYNCTAVNRIGQES 245

Qy     198 LDFHLIVKDKTVGVP----IEPIIAACVVIF 224
```

```
          |:| |:  | |   |    :|| :   | |
Db    246 LEFILVQAD-TPSSPSIDQVEPYSSTAQVQF 275


RESULT   4
P91255
ID   P91255      PRELIMINARY;      PRT;  2762 AA.
AC   P91255;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 306.5 KDA PROTEIN.
GN   F12F3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton B., Wohldmann P.;
RT   "The sequence of C. elegans cosmid F12F3.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80022; AAC25886.2; -
KW   Hypothetical protein.
SQ   SEQUENCE   2762 AA;  306525 MW;  C964CCEB6D322084 CRC64;


  Query Match           13.4%;  Score 173.5;  DB 5;  Length 2762;
  Best Local Similarity   24.1%;  Pred. No. 4.6e-06;
  Matches   60;  Conservative   36;  Mismatches  100;  Indels   53;  Gaps    9;

Qy       4 GRFLLLVILFLPREMTSSVLTVNGKTENYILDTTP----------------------GS 40
           | |  |:   | :   |  :|:  |  :  | :||                      |
Db    1562 GTFNCLIENELGQASASCQVTIFNKPAS--LQSTPDHSLERNLVPTLQKALNNESAQAGQ 1619

Qy      41 QASLICAVQNHTREEELLWYREE------GRVDLKSGNKINSSSVCVSSISENDNGISFT 94
           |  | | : : : |  : |::::      || :| |  | |   ||  ::   | |  :
Db    1620 QIMLTCRISSRS-ESTVAWFKDDERIESAGRYFLSSDKKSNHKLVC-HAVQSQDTG-KYR 1676

Qy      95 C----RLGRDQS---VSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMW 147
           |    : |  :|   |:|  |     |   :     |   | |: | | |: :|  :: |
Db    1677 CVVTNKYGYAESECNVAVEDVTKFIAPSFSATLSDSTAILGHNITLECKVEGSPAPEVSW 1736

Qy     148 YKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKS-----------SLKT 195
            |:   :   :   | |  : : :|||:| |  | | | | | |             ||
Db    1737 TKDGERISTTRRIRQTQDENGNCKLSISKAESDDMGVYVCSATSVAGVDSTSSMVMIAKT 1796

Qy     196 ESLDFHLIV 204
             | ||::
Db    1797 TGTDSHLVI 1805


RESULT   5
Q9ERC8
ID   Q9ERC8      PRELIMINARY;      PRT;  2013 AA.
AC   Q9ERC8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DOWN SYNDROME CELL ADHESION MOLECULE.
GN   DSCAM.
OS   Mus musculus (Mouse).
OC   Eukaryota, Metazoa, Chordata, Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=BRAIN;
RA   Barlow G.M., Lyons G.E., Korenberg J.R.;
RT   "Down syndrome cell adhesion molecule. a highly conserved mouse
RT   homolog, dscam, is expressed in the neonatal and adult mouse brain.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR OUT-BRED;
RX   MEDLINE=21134302; PubMed=11237714;
RA   Agarwala K.L., Ganesh S., Amano K., Suzuki T., Yamakawa K.;
RT   "DSCAM, a highly conserved gene in mammals, expressed in
RT   differentiating mouse brain.";
RL   Biochem. Biophys. Res. Commun. 281:697-705(2001).
DR   EMBL; AF315558; AAG28796.1; -.
DR   EMBL; AY005483; AAF99440.1; -.
DR   MGD; MGI:1196281; Dscam.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 8.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00409; IG; 9.
DR   SMART; SM00408; IGc2; 9.
DR   SMART; SM00410; IG_like; 6.
KW   Immunoglobulin domain.
SQ   SEQUENCE   2013 AA;  222268 MW;  1F4AF0794788EBA77 CRC64;


  Query Match           13.3%;  Score 172;  DB 11;  Length 2013;
  Best Local Similarity   29.8%;  Pred. No. 4.2e-06;
  Matches   57;  Conservative   33;  Mismatches   67;  Indels   34;  Gaps   10;

Qy      15 PREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKSGNKI 74
           ||:: |||                 ||| || |:|  :  ::|| |||  | : |  |  :
Db    320 PRKVKSSV----------------GSQVSLSCSVTGN-EDQELSWYR-NGEI-LNPGKNV 360

Qy      75 -----NSSSVCVSSISENDNGISFTCRLGRDQSVS---VSVVLNVTFPPLLSGNDFQTVE 126
                | ::: :  : ::| | :: | : ::|:  :   | |||    | ::|    : |
Db    361 KITGLNHANLIMDHMVKSDGG-AYQCFVRNDKLSAQDYVQVVLEDGTPKIISAFSEKVVS 419

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQ--TSES---FQISITKVEKPD 181
           | |||||  |   : |  :   : |: | |:| |  |||      |:|:  :  |
Db    420 PAEPVSLVCNVKGTPLPTVTWTLDDDPI-LKGSGHRISQMITSEGNVVSYLNISSSQVRD 478

Qy     182 NGTYSCIAKSS 192
             | | | | :|
Db    479 GGVYRCTANNS 489


RESULT   6
Q07153
ID   Q07153      PRELIMINARY;      PRT;   946 AA.
AC   Q07153;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RECEPTOR TYROSINE KINASE.
OS   Torpedo californica (Pacific electric ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Torpediniformes, Torpedinoidei, Torpedinidae, Torpedo.
```

```
OX   NCBI_TaxID=7787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ELECTRIC ORGAN, AND MUSCLE;
RX   MEDLINE=93219391; PubMed=8385349;
RA   Jennings C.G.B., Dyer S.M., Burden S.J.;
RT   "Muscle-specific trk related receptor with a kringle domain defines a
RT   distinct class of receptor tyrosine kinases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:2895-2899(1993).
DR   EMBL; L11311; AAA49285.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000024; Fz_domain.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF01392; Fz; 1.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00051; kringle; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR00109; TYRKINASE.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00130; KR; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50038; FZ; 1.
DR   PROSITE; PS00021; KRINGLE_1; FALSE_NEG.
DR   PROSITE; PS50070; KRINGLE_2; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Immunoglobulin domain; Kinase; Transferase;
KW   Tyrosine-protein kinase.
SQ   SEQUENCE   946 AA;  105892 MW;  81A9131DB7479D01 CRC64;


  Query Match             13.2%;  Score 171.5;  DB 13;  Length 946;
  Best Local Similarity   25.1%;  Pred. No. 1.8e-06;
  Matches   63;  Conservative   38;  Mismatches   95;  Indels   55;  Gaps   10;

Qy       7 LLLVIL         FLPR  EMTSSVLTVNCKTENYTLMTTPGSQASLICAVQNHTREE 55
           ||:: |         ||:  ::|| : ||:   |         :|| :||| ::
Db      10 LLMIFLVTTGGSAMGILPKAPQITSPLETVDALVE---------FEASFMCAVDSYP-AA 59

Qy      56 ELLWYREEGRV---DLKSGNKINSSSVCVSSISENDNGI-SFTCRLGRDQSVSVSVVLNV 111
           |: | |    :   | :   | |   : : |: : |||:   |   |   |       | |
Db      60 EITWTRNNIPIRPEDIRYSIKENGQTLTILSVEDIDNGVYGCIANNGMGSSAQSCGALQV 119

Qy     112 TFPP  LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSES 169
             |  :    | : :  || | | |:   ||:  : |:|: : |  :: |  :    ||
Db     120 KMKPKIIRPPIDVRAL-LGSKVVLPCSIMGNPEPAISWFKDEIALKND-PRISV---IES 175

Qy     170 FQLSITKVEKPDNGTYSCIAKSSLKTE----------------------SLDFHLIVK 205
           | |  |:  | | | |:|::||  |                          |    :|::
Db     176 GNLRIRNVQLEDAGKYRCLARNSLGFEYSRSAALEVQVSARIVKAPTSQNVSYGSEVILQ 235

Qy     206 DKTVGVPIEPI 216
           |  | ||  |
Db     236 CKATGFPIPTI 246


RESULT   7
Q9V6C2
ID   Q9V6C2      PRELIMINARY;      PRT;   359 AA.
AC   Q9V6C2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LAC PROTEIN (LD33460P).
GN   LAC OR CG12369.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y. H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z. Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R. F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003822; AAF58506.1; -.
DR   EMBL; AY051829; AAK93253.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0010238; Lac.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   359 AA;  39939 MW;  12F513E2B9C156F4 CRC64;


  Query Match             13.1%;  Score 169.5;  DB 5;  Length 359;
  Best Local Similarity   27.1%;  Pred. No. 8.4e-07;
  Matches   56;  Conservative   42;  Mismatches   78;  Indels   31;  Gaps   10;
```

```
Qy       8 LLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-EEGRV 66
           ||: :|: : :     |::  |:  | |   |      |:|| : :|  :|: : :   |
Db      14 LLLAIFVQQTLAQRTPTISYITQEQIKDI--GGTVEFDCSVQ-YAKEYNVLFLKTDSDPV 70

Qy      67 DLKSGNKI------------NSSS--VCVSSISENDNGISFTCR--LGRDQSVSVSVVL 109
             | :|: :              |||:  : :  | | | | ::||:  :     ||  | |
Db      71 FLSTGSTLVIKDSRFSLRYDPNSSTYKLQIKDIQETDAG-TYTCQVVISTVHKVSAEVKL 129

Qy     110 NVTFPPLLSGNDFQTV--EEGSNVKLVCNVKANPQAQMMWYK-NSSLLDLEKSKHQIQQT 166
           .|  ||..| |  |.|   ||| |.. |     |  .| . |...|  . .      ;
Db     130 SVRRPPVISDNSTQSVVASEGSEVQMECYASGYPTPTITWRRENNAILPTDSA------T 183

Qy     167 SESFQLSITKVEKPDNGTYSCIAKSSL 193
               | |  |:| | ||| |:| : :
Db     184 YVGNTLRIKSVKKEDRGTYYCVADNGV 210


RESULT   8
Q9NYZ4
ID   Q9NYZ4      PRELIMINARY;      PRT;   431 AA.
AC   Q9NYZ4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIGLEC SAF2.
GN   SAF2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kikly K.K., Bochner B.S., Freeman S.D., Tan K.B., Gallagher K.T.,
RA   D'Alessio K.J., Holmes S.D., Abrahamson J.A., Erickson-Miller C.L.,
RA   Murdock P.R., Tachimoto H., Schleimer R.P., White J.R.;
RT   "Identification of SAF-2, a novel siglec expressed on eosinophils,
RT   mast cells and basophils.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF223403; AAF34702.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   431 AA;  46433 MW;  49B91B84CE67B118 CRC64;


  Query Match             13.1%;  Score 169.5;  DB 4;  Length 431;
  Best Local Similarity   23.6%;  Pred. No. 1.1e-06;
  Matches   63;  Conservative   53;  Mismatches   92;  Indels   59;  Gaps   11;

Qy      16 REMTSSVLTVNGKTENYILDT-TPGSQASLICAVQNHTREEELLWYREEGRVDLKS--GN 72
           ::::  |  :  : :  || |   |   :| |:|         |  ::|   : | |
Db     144 KQLSVFVTALTHRPDILILGTLESGHSRNLTCSVP---------WACKQGTPPMISWIGA 194

Qy      73 KINS-------SSVCVSSISENDNGISFTCRL---GRDQSVSVSVVLNVTFPP------- 115
            ::|       |||   :    |:| | ||::   |   : : :| |:|::||
Db     195 SVSSPGPTTARSSVLTLTPKPQDHGTSLTCQVTLPGTGVTTTSTVRLDVSYPPWNLTMTV 254

Qy     116 ---------LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQT 166
                    |      :| || :::||| | :|| |:: | : |   | |  ||     :
Db     255 FQGDATASTALGNGSSLSVLEGQSLRLVCAVNSNPPARLSWTPGS--ITICPSP-----S 307

Qy     167 SESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPIIAACV--- 221
           |    | : :|   | | ::| |:::  :: :   | ::::  |   |:  :  | |
Db     308 SNPGLLELPRVHVRDEGEFTCRAQNAQGSQHISLSLSLQNEGTGTSRPVSQVTLAAVGGA 367

Qy     222 ----VIFLTLCFGLIARRKKIMKLCMK 244


Db     368 GATALAFLSFCIIFI-----IVRSCRK 389


RESULT   9
Q921P2
ID   Q921P2      PRELIMINARY;      PRT;   605 AA.
AC   Q921P2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROTEIN FOR MGC:12077).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011310; AAH11310.1; -.
SQ   SEQUENCE   605 AA;  67353 MW;  DBD98C64D245A867 CRC64;


  Query Match             12.8%;  Score 166;  DB 11;  Length 605;
  Best Local Similarity   28.0%;  Pred. No. 3.2e-06;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      41 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 86
           | :| |               : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEKTL----DGHMVVRSHARV--SSLTLKSIQY 378

Qy      87 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 142
             | |  : |     :|:|   | |: | | : | | |        ||: | : | | | |
Db     379 TDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     143 AQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSIIKVEKPDNGTYSCIAKSSLKTESLDFH 201
           | : |:::  ||     |  :| |  :  | :|   : | | |:| | : :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     202 LIVKDKTVGVP----IEPIIAACVVIF 224
           |:  | |   |    :||  :   | |
Db     495 LVQAD-TPSSPSIDRVEPYSSIAQVQF 520


RESULT  10
Q24273
ID   Q24273      PRELIMINARY;      PRT;  1011 AA.
AC   Q24273;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROMUSCULIN.
GN   NRM OR CG8779.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94000831; PubMed=8398154;
RA   Kania A., Han P.L., Kim Y.T., Bellen H.;
RT   "Neuromusculin, a Drosophila gene expressed in peripheral neuronal
RT   precursors and muscles, encodes a cell adhesion molecule.";
RL   Neuron 11:673-687(1993).
DR   EMBL; L23146; AAA03750.1; -.
DR   FlyBase; FBgn0005629; nrm.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 9.
DR   SMART; SM00409; IG; 4.
```

```
DR   SMART; SM00410; IG_like; 2.
KW   Cell adhesion.
SQ   SEQUENCE   1011 AA;  113125 MW;  AFD9A2A015D3AE63 CRC64;


  Query Match             12.4%;  Score 161;  DB 5;  Length 1011;
  Best Local Similarity   24.3%;  Pred. No. 1.6e-05;
  Matches   46;  Conservative   34;  Mismatches   81;  Indels   28;  Gaps    6;

Qy      23 LTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-----EEGRVDL--------- 68
           : :||   :    |  ||:  | | :         | ||        |  ::
Db     265 IDINGVKHH----TVQGSKVVLTCDIHGARPAVNLTWYNTTTIISSGENEITEVRSKSLE 320

Qy      69 KSGNKINSSSVCVSSISENDNGISFTC-------RLGRDQSVSVSVVLNVTFPPLLSGND 121
           ||    :: |  : : :  :|   | |        :: |:: :| :: | | :|||::  :
Db     321 KSDGTFHIQSELIFNATREENDKVFRCEAENIVIQINREKPISSALTLEVLYPPVVKVSP 380

Qy     122 FQTVEEGSNVKLV-CNVKANPQA--QMMWYKNSSLLDLEKSRHQIQQTSESPQLSITKVE 178
                 | : |: |   ||| :  |: ||:|  |::  : |    ||:  | |   |
Db     381 SAITANTSEIVLLNCEYFANPASLTQVEWYRNDILVNVNDTTHYKGGNSENVALVIKSTE 440

Qy     179 KPDNGTYSC 187
           | | | |||
Db     441 KEDIGNYSC 449


RESULT  11
Q9VNP2
ID   Q9VNP2      PRELIMINARY;      PRT;  1079 AA.
AC   Q9VNP2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NRM PROTEIN.
GN   NRM OR CG8779.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H. J., Andrews Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z. Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003599; AAF51887.1; -.
DR   FlyBase; FBgn0005629; nrm.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 9.
DR   SMART; SM00409; IG; 4.
DR   SMART; SM00410; IG_like; 2.
SQ   SEQUENCE   1079 AA;  121454 MW;  C765558221AC127B CRC64;


  Query Match             12.4%;  Score 161;  DB 5;  Length 1079;
  Best Local Similarity   24.3%;  Pred. No. 1.8e-05;
  Matches   46;  Conservative   34;  Mismatches   81;  Indels   28;  Gaps    6;

Qy      23 LTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-----EEGRVDL--------- 68
           : :||   :    |  ||:  | | :         | ||        |  ::
Db     265 IDINGVKHH----TVQGSKVVLTCDIHGARPAVNLTWYNTTTIISSGENEITEVRSKSLE 320

Qy      69 KSGNKINSSSVCVSSISENDNGISFTC-------RLGRDQSVSVSVVLNVTFPPLLSGND 121
           ||    :: |  : : :  :|   | |        :: |:: :| :: | | :|||::  :
Db     321 KSDGTFHIQSELIFNATREENDKVFRCEAENIVIQINREKPISSALTLEVLYPPVVKVSP 380

Qy     122 FQTVEEGSNVKLV-CNVKANPQA--QMMWYKNSSLLDLEKSRHQIQQTSESPQLSITKVE 178
                 | : |: |   ||| :  |: ||:|  |::  : |    ||:  | |   |
Db     381 SAITANTSEIVLLNCEYFANPASLTQVEWYRNDILVNVNDTTHYKGGNSENVALVIKSTE 440

Qy     179 KPDNGTYSC 187
           | | | |||
Db     441 KEDIGNYSC 449


RESULT  12
Q95LH0
ID   Q95LH0      PRELIMINARY;      PRT;  597 AA.
AC   Q95LH0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIALIC ACID-BINDING LECTIN SIGLEC-11.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21523976; PubMed=11546777;
RA   Angata T., Varki N.M., Varki A.;
RT   "A second uniquely human mutation affecting sialic acid biology.";
RL   J. Biol. Chem. 276:40282-40287(2001).
DR   EMBL; AF293372; AAL09302.1; -.
KW   Lectin.
SQ   SEQUENCE   597 AA;  65075 MW;  74C13CFA9CDB5BA5 CRC64;


  Query Match             12.4%;  Score 159;  DB 6;  Length 597;
  Best Local Similarity   23.2%;  Pred. No. 1.3e-05;
  Matches   63;  Conservative   45;  Mismatches   85;  Indels   78;  Gaps   12;

Qy      31 NYILD--------------TTPGSQAS-----LICAVQNHTREEELLWYREEGRVDLKS 70
           ||| |              : ||:  |     | |:|          |  |:|     :
Db     260 NYIYDKLSVHVTALTHLPTFSIPGTLESGHPRNLTCSVP---------WACEQGTPPTIT 310

Qy      71 --GNKINSSSVCVSSIS-------ENDNGISFTCRL---GRDQSVSVSVVLNVTFPP--- 115
```

```
Db    311 WMGASVSSLEPTISRSSMLSLIPKPQDHGTSLTCQVTLPGAGVTTTRAVRLNISYPPQNL 370

Qy    116 ------------LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQ 162
                      |      :| || :: ||| | :|| |:: |   |  | |  |
Db    371 TMTVFQGDGTASTTLRNGSALSVLEGQSLHLVCAVDSNPPARLSWTWGS--LTLSPS--- 425

Qy    163 IQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPI---- 216
            |:|    | : :|   | | ::| |:: | :: :   | :::: |   ||  :
Db    426 --QSSNLGVLELPRVHVKDEGEFTCRAQNPLGSQHISLSLSLQNEYTGKMRPISGVTLGA 483

Qy    217 ---IAACVVIFLTLCFGLIARRKKIMKLCMK 244
                |  ::||: |   :     ::: | |
Db    484 VGGAGATALVFLSFCIIFV-----VVRSCRK 509


RESULT  13
O55124
ID   O55124      PRELIMINARY;      PRT;  1463 AA.
AC   O55124;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   M-PROTEIN.
GN   MYOM2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98234547; PubMed=9570952;
RA   Steiner F., Weber K., Fuerst D.O.;
RT   "Structure and expression of the gene encoding murine M-protein, a
RT   sarcomere-specific member of the immunoglobulin superfamily.";
RL   Genomics 49:83-95(1998).
DR   EMBL; AJ001038; CAA04492.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:1328358; Myom2.
DR   InterPro; IPR003962; FnIII_repeat
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 4.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1463 AA;  164518 MW;  8C6551ED90F0B288 CRC64;


  Query Match             12.2%;  Score 158;  DB 11;  Length 1463;
  Best Local Similarity   25.1%;  Pred. No. 4.7e-05;
  Matches   60;  Conservative   36;  Mismatches   97;  Indels   46;  Gaps   10;

Qy     21 SVLTVNGKT-ENYIL--DTTPGSQAS-----------LICAVQNHTREEELLWYREEGR 65
          ||| | ||  |: ||      |: ||           | | ::  | | :: || :| :
Db   1211 SVLEVGGKVYEDMILAMSRVCGASASPLKVLCTPEGIRLQCFMKYFTEEMKVSWYHKEAK 1270

Qy     66 VDLKSGNKINSSS----VCVSSISENDNG-ISFTCRLGRD-------------------- 100
          :      :|  |     : :   :| | |  :|    |:|
Db   1271 ISSSEHMRIGGSEEMAWLQICEPTEKDKGKYTFEIFDGKDSHQRSLDLSGQAFDEAYAEF 1330

Qy    101 QSVSVSVVLNVTFPPLLSG-NDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKS 159
          | :  :          :: |  |  |. || . | | |  ||  ...|.||    |.| |
Db   1331 QQLKAAAFAEKNRGKVIGGLPDVVTIMEGKTLNLTCTVFGNPDPEVVWFKNDK--DIELS 1388

Qy    160 RHQIQQTSESFQLSIT--KVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGVP-IEP 215
          | : :  :|  :|:|   |   |:| ||   |:    | :|  : |      :| | |
Db   1389 EHFLVKMEQSKYVSLTIQGVTAEDSGKYSINVKNKYGGEKIDVTVSVYKHGEKIPDISP 1447
```

```
RESULT  14
Q94155
ID   Q94155      PRELIMINARY;      PRT;  1415 AA.
AC   Q94155;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNC-40 (T19B4.6 PROTEIN).
GN   UNC-40 OR T19B4.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL;
RA   Chan S.S.Y., Zheng H., Su M.W., Wilk R., Killeen M.T., Hedgecock E.M.,
RA   Culotti J.G.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1248-1415 FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [3]
RP   SEQUENCE OF 1248-1415 FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Gattung S.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1248-1415 FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U70618; AAB17088.1; -.
DR   EMBL; U80438; AAB37634.1; -.
DR   HSSP; P40189; 1BQU.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 1.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1415 AA;  154436 MW;  B2D90F76F9FEB6A9 CRC64;


  Query Match             12.2%;  Score 157.5;  DB 5;  Length 1415;
  Best Local Similarity   27.8%;  Pred. No. 5e-05;
  Matches   58;  Conservative   35;  Mismatches   83;  Indels   33;  Gaps   10;

Qy     16 REMTSSVLTVNGKT--ENYILDTTP-------GSQASLICAVQNHTREEELLWYREEGR- 65
          |   :: |||  :|     :  |||        | :  | | | :  |  :: | ::  :
Db    227 RSSQTARLTVTTETVSNELVFITTPRLQVVEQGDEFLLECLVASLIR-PQVRWLKDSRQ1 285
```

```
Qy      66 -VDLKSGNKINSSSVCVSSISENDNGISFTCPL-GRDQSVSVSVVLNVTFPPLLSGNDFQ 123
           ||      ::  ||: ||  |  | |: :|||    | |:  :| : |  || ::
Db     286 IVDGVRIRRVGVSSILVSRASIEDTGL-YTCKASNNDDSIDRAVSVEVKAPPRIIIRPII 344

Qy     124 TVE-EGSNVKLVC-NVKANPQAQMMWYKN------SSLLDLEKSRHQIQQTSESFQLSII 175
           |  | ::|:| |    | |:|:: ||||      |    :| :|           | |
Db     345 KVAVETADVELECGTAAARPEARVNWYKNGEAIIGSEYFVIEPNR----------LRIL 393

Qy     176 KVEKPDNGTYSCIAKSSLKTESLDFHLIV 204
           | : |   | |||:: : :|     |:|
Db     394 GVVRADQAIYQCIAENDVGSEQASAQLLV 422


RESULT  15
Q9VP08
ID   Q9VP08      PRELIMINARY;      PRT;   467 AA.
AC   Q9VP08;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG7166 PROTEIN.
GN   CG7166.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pteryqota; Neoptera; Endopteryqota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003595; AAF51754.1; -.
DR   FlyBase; FBgn0037107; CG7166.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   467 AA;  53023 MW;  5D0DCCF9CD3BBB7E CRC64;


  Query Match             12.1%;  Score 157;  DB 5;  Length 467;
  Best Local Similarity   26.5%;  Pred. No. 1.4e-05;
  Matches   48;  Conservative   37;  Mismatches   72;  Indels   24;  Gaps    8;

Qy      24 TVNCKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKSG--NKINSSSVCV 81
           |:     |  :    ||  :| |    :     : |:::    |: ||  :  :|||: :
Db     136 TLRALPHNGQVTARKGSTVTLECKASGNP-VPTIEWFKK----DVFSGPIHLSDSSTLIL 190

Qy      82 SSISFNDNGISFTCPL--GRDQSVSVSVVLNVTFPPLLSGNDFQTVE-------EGSNVK 132
           ::  :  | :: |    |    ||: : | :  || :       |||        || :|:
Db     191 ENVDRHHAG-TYQCSADNGVKDRVSMDIQLTILSPPEI------TVEKSWVHASEGYDVE 243

Qy     133 LVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSIIKVEKPDNGTYSCIAKSS 192
           ||| |  :  ::|:|||:|| |||    |  :    : : | |   :  | | |||:| ::
Db     244 LVCIVHGDVNSEMLWYQNSFLLD-AIDRKSMYFRDDRYSLIIRNFQPTDFGNYSCVALNA 302

Qy     193 L 193
           |
Db     303 L 303


Search completed: September 17, 2002, 13:23:22
Job time: 804 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:23:22 ; Search time 102.65 Seconds
(without alignments)
441.546 Million cell updates/sec

Title: US-09-863-823-6
Perfect score: 1336
Sequence: 1 MAWKSSVIMQMGRFLLLVIL.....RKKIMKLCMKLKDPHSETAL 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap.*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	973.5	72.9	261	11	Q9D7L8	Q9d718 mus musculu
2	922.5	69.0	253	11	Q9D8H2	Q9d8h2 mus musculu
3	175	13.1	604	4	Q96CJ3	Q96cj3 homo sapien
4	173.5	13.0	2762	5	P91255	P91255 caenorhabdi
5	173	12.9	359	5	Q9V6C2	Q9v6c2 drosophila
6	172	12.9	2013	11	Q9ERC8	Q9erc8 mus musculu
7	171.5	12.8	946	13	Q07153	Q07153 torpedo cal
8	169.5	12.7	431	4	Q9NYZ4	Q9nyz4 homo sapien
9	166	12.4	605	11	Q921P2	Q921p2 mus musculu
10	161	12.1	1011	5	Q24273	Q24273 drosophila
11	161	12.1	1079	5	Q9VNP2	Q9vnp2 drosophila
12	159	11.9	597	6	Q95LH0	Q95lh0 pan troglod
13	158	11.8	1463	11	O55124	O55124 mus musculu
14	157.5	11.8	1415	5	Q94155	Q94155 caenorhabdi
15	157	11.8	467	5	Q9VP08	Q9vp08 drosophila
16	156	11.7	1173	5	P91456	P91456 caenorhabdi
17	155	11.6	595	4	Q96PQ1	Q96pq1 homo sapien
18	154.5	11.6	1513	13	Q90Z70	Q90z70 brachydanio
19	154	11.5	509	11	Q9EQY5	Q9eqy5 mus musculu
20	154	11.5	509	11	Q91YK7	Q91yk7 mus musculu
21	154	11.5	603	5	Q9NKF5	Q9nkf5 drosophila
22	154	11.5	868	11	Q62838	Q62838 rattus norv
23	153.5	11.5	850	4	Q9ULF7	Q9ulf7 homo sapien
24	153.5	11.5	1344	11	Q922I4	Q922i4 mus musculu
25	153	11.5	467	4	Q9Y286	Q9y286 homo sapien
26	152.5	11.4	5198	5	O76518	O76518 caenorhabdi
27	152	11.4	869	4	O15146	O15146 homo sapien
28	149	11.2	860	11	Q61005	Q61005 mus musculu
29	149	11.2	868	11	Q61006	Q61006 mus musculu
30	149	11.2	2037	5	Q9V1S8	Q9v1s8 drosophila
31	148.5	11.1	4370	4	Q9H3V5	Q9h3v5 homo sapien
32	147.5	11.0	496	5	Q94224	Q94224 caenorhabdi
33	147	11.0	334	7	O02870	O02870 gallus gall
34	147	11.0	1445	11	Q63155	Q63155 rattus norv
35	146.5	11.0	528	5	P91670	P91670 drosophila
36	146.5	11.0	545	5	Q9VCT4	Q9vct4 drosophila
37	146.5	11.0	773	4	Q9NSW7	Q9nsw7 homo sapien
38	146	10.9	7962	4	Q10465	Q10465 homo sapien
39	145.5	10.9	1215	5	Q9V787	Q9v787 drosophila
40	145	10.9	951	5	Q19128	Q19128 caenorhabdi
41	144.5	10.8	350	7	O02869	O02869 gallus gall
42	144.5	10.8	693	4	Q9UPU1	Q9upu1 homo sapien
43	144.5	10.8	1482	5	Q9V4Y0	Q9v4y0 drosophila
44	144.5	10.8	2016	5	Q9V4J9	Q9v4j9 drosophila
45	144.5	10.8	6620	4	Q96AA2	Q96aa2 homo sapien

ALIGNMENTS

RESULT 1
Q9D7L8
ID Q9D7L8 PRELIMINARY; PRT; 261 AA.
AC Q9D7L8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 2010002A20RIK PROTEIN.
GN 2010002A20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK009118; BAB26083.1; -.
DR MGD; MGI:1913851; 2010002A20Rik.

```
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00409; IG; 2.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   261 AA;  29050 MW;  1FF1FC02D27E83CD CRC64;


  Query Match             72.9%;  Score 973.5;  DB 11;  Length 261;
  Best Local Similarity   73.8%;  Pred. No. 6.4e-77;
  Matches  192;  Conservative   28;  Mismatches   41;  Indels    1;  Gaps    1;

Qy       1 MAWKSSVIMQMGRFLLLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHT 60
           | ||  :  :|  : ||||:| ||:  ||||||||||:||||||||  | |||| |||||||
Db       1 MVWKITGPLQACQ-LLLVVLSLPQGRTSSVLTVNGKTENYILDTQHGVQASLECAVQNHT 59

Qy      61 REEELLWYREEGRVDLKSGNKINSSSVCVSSISENDNGISFTCRLGRDQSVSVSVVLNVT 120
            :|||||||||:| ||||:|||||  ||||||  |:|:|||:  |||:|  |||:|||:||||||
Db      60 EDEELLWYREDGIVDLKNGNKINISSVCVSPINESDNGVRFTCKLQRDQTVSVTVVLNVT 119

Qy     121 FPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQL 180
           ||||||||  ||||||  |:|  |||||:|||||||||||:| |  |||  ||||  ||  |||||
Db     120 FPPLLSGNGFQTVEENSDVSLVCNVKSNPQAQMMWYKNNSALVLEKGRHQIHQTRESFQL 179

Qy     181 SITKVEKPDNGTYSCIAKSSLKTESLDFHLLVKDKTVGVPTEPTIAACVVIFLTLCFGLI 240
           ||||||:|  ||||||||||  ||||  |::||||:||||    :|  ||||||||||: ||: | |
Db     180 SITKVKKSDNGTYSCIASSSLKMETMDFHLLVKDKVFVMPAEPTIAACVVVVLTMAFALF 239

Qy     241 ARRKKIMKLCMKDKDPHSETAL 262
           :|||:|||||  |  ||:|||||
Db     240 SRRKRIMKLCGKKNDPNSETAL 261


RESULT   2
Q9D8H2
ID   Q9D8H2      PRELIMINARY;      PRT;   253 AA.
AC   Q9D8H2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   2010002A20Rik PROTEIN.
GN   2010002A20RIK
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Matchinkeni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schonbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection";
RL   Nature 409:685-690(2001).
DR   EMBL; AK008028; BAB25419.1; -.
DR   MGD; MGI:1913851; 2010002A20Rik.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00409; IG; 2.
DR   SMART; SM00408; IGc2; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   253 AA;  28372 MW;  BCA6EC8C8D7167D1 CRC64;


  Query Match             69.0%;  Score 922.5;  DB 11;  Length 253;
  Best Local Similarity   73.4%;  Pred. No. 1.7e-72;
  Matches  182;  Conservative   27;  Mismatches   38;  Indels    1;  Gaps    1;

Qy       1 MAWKSSVIMQMGRFLLLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHT 60
           | ||  :  :|  : ||||:| ||:  ||||||||||:||||||||  | |||| |||||||
Db       1 MVWKITGPLQACQ-LLLVVLSLPQGRTSSVLTVNGRTENYILDTQHGVQASLECAVQNHT 59

Qy      61 REEELLWYREEGRVDLKSGNKINSSSVCVSSISENDNGISFTCRLGRDQSVSVSVVLNVT 120
            :|||||||||:| ||||:|||||  ||||||  |:|:|||:  |||:|  |||:|||:||||||
Db      60 EDEELLWYREDGIVDLKNGNKINISSVCVSPINESDNGVRFTCKLQRDQTVSVTVVLNVT 119

Qy     121 FPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQL 180
           ||||||||  ||||||  |:|  |||||:|||||||||||:| |  |||  ||||  ||  |||||
Db     120 FPPLLSGNGFQTVEENSDVSLVCNVKSNPQAQMMWYKNNSALVLEKGRHQIHQTRESFQL 179

Qy     181 SITKVEKPDNGTYSCIAKSSLKTESLDFHLLVKDKTVGVPTEPTIAACVVIFLTLCFGLI 240
           ||||||:|  ||||||||||  ||||  |::||||:||||    :|  ||||||||||: ||: | |
Db     180 SITKVKKSDNGTYSCIASSSLKMETMDFHLLVKDKVFVMPAEPTIAACVVVVLTMAFALF 239

Qy     241 ARRKKIMK 248
           :|||:|||
Db     240 SRRKRIMK 247


RESULT   3
Q96CJ3
ID   Q96CJ3      PRELIMINARY;      PRT;   604 AA.
AC   Q96CJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO NEURAL CELL ADHESION MOLECULE 1 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY, AND RENAL CELL ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC014205; AAH14205.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   604 AA;  66271 MW;  B3B791A4D033D546 CRC64;


  Query Match             13.1%;  Score 175;  DB 4;  Length 604;
  Best Local Similarity   28.4%;  Pred. No. 5.8e-07;
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      49 QASLIC----------AVQNHTREEELLWYREEGRVDLKSGNKINS----SSVCVS 90
           | :| |            : :| : ||:  | | | :  |    :  |    ||: :
Db      70 QVILICEASGDPIPSITWRKISIENISSEEKASWTLRPEKQEITLAGHMVVKSHARVSSLTLK 129

Qy      91 SISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 146
           ||   | |  : |      :|:|  | |: | | : | | |       ||: | : | |
Db     130 SIQYIDAGEYICIASNTIGQD---SQSMYLEVQYAPKLQGPVAVYIWEGNQVNITCEVF 185

Qy     147 ANPQAQMMWYKNSSLIDLEK-SPHQIQQTSFSFQLSITKVEKPDNGTYSCIAKSSLKTES 205
```

```
          | | | : |::: ||     |  :|  |  :  | :|   : | | |:| | : :  ||
Db    186 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 245

Qy    206 LDFHLIVKDKTVGVP----IEPIIAACVVIF 232
          |:| |:  | |   |    :||  :   | |
Db    246 LEFILVQAD-TPSSPSIDQVEPYSSTAQVQF 275


RESULT   4
P91255
ID   P91255      PRELIMINARY;      PRT;  2762 AA.
AC   P91255;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 306.5 KDA PROTEIN.
GN   F12F3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton B., Wohldmann P.;
RT   "The sequence of C. elegans cosmid F12F3.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL: U80022: AAC25886.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2762 AA;  306525 MW;  C964CCEB6D322084 CRC64;


  Query Match             13.0%;  Score 173.5;  DB 5;  Length 2762;
  Best Local Similarity   24.1%;  Pred. No. 5e-06;
  Matches   60;  Conservative   36;  Mismatches  100,  Indels   53,  Gaps    9,

Qy     12 GRFLLLVILFLPREMTSSVLTVNGKTENYILDTTP-----------------------GS 48
          | |  |:   | :   |  :|:  |  :  | :||                       |
Db   1562 GTFNCLIENELGQASASCQVTIFNKPAS--LQSTPDHSLERNLVPTLQKALNNESAQAGQ 1619

Qy     49 QASLICAVQNHTREEELLWYREE------GRVDLKSGNKINSSSVCVSSISENDNGISFT 102
          |  | | : : :  : |::::      || :| |  | |   ||  ::   | |  :
Db   1620 QIMLTCRISSRS-ESTVAWFKDDEPIESAGRYFLSSDKKSNHKLVC-HAVQSQDTG-KYR 1676

Qy    103 C----RLGRDQS---VSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMW 155
          |    : |  :|   |:|  |     |   :     |   | |: | | |: :|  :: |
Db   1677 CVVTNKYGYAESECNVAVEDVTKFIAPSFSATLSDSTAILGHNITLECKVEGSPAPEVSW 1736

Qy    156 YKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKS------------SLKT 203
           |:   :   :   | | : : :|||:| |  | | | | | |              ||
Db   1737 TKDGERISTTRRIRQTQDENGNCKLSISKAESDDMGVYVCSATSVAGVDSTSSMVMIAKT 1796

Qy    204 ESLDFHLIV 212
             | ||::
Db   1797 TGTDSHLVI 1805


RESULT   5
Q9V6C2
ID   Q9V6C2      PRELIMINARY;      PRT;  359 AA.
AC   Q9V6C2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LAC PROTEIN (LD33460P).
GN   LAC OR CG12369.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003822; AAF58506.1; -.
DR   EMBL; AY051829; AAK93253.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0010238; Lac.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
```

```
KW   Immunoglobulin domain.
SQ   SEQUENCE   459 AA;  49949 MW;  12F513E2B9C156F4 CRC64;


  Query Match             12.8%;  Score 172;  DB 5;  Length 459;
  Best Local Similarity   27.3%;  Pred. No. 4.6e-07;
  Matches   60;  Conservative   43;  Mismatches   85;  Indels   32;  Gaps   11;

Qy       3 WKSSVIMQMGREELLVILFLVSLGVNGKTENYTLDTIPGSQASLPCAVQNHTRE 62
           |: |:   :   ||| | |: : :     |::  |:  | |   |       |:|| : :|
Db       2 WRPSTSNCVWSTLLLAI-FVQQILAQPIPLTSYTEQFQTKDI---QSTVEFPCSVQ-YAKE 57

Qy      63 EELLWYR-EEGRVDLKSGNKI-------------NSss--VCVSSISENDNGISFTCR-- 104
             :|: : :   | | :|: :               |||:  : :  | | | | ::||: 
Db      58 YNVLFLKTDSDPVFLSTGSTIVIKDSREFSLPYDPNSSTYKTQTKDIQETIDAG-TYPCQVV 116

Qy     105 LGRDQSVSVSVVINVTFPPLLSGNDPQTV--FFGSNVKLVCNVKANPQAQMMWYK-NSSI 161
           :     ||  | |:|  ||::| |  |:|   ||| |:: |     |   : | : |:::
Db     117 ISTVHKVSAEVKLSVRRPPVISDNSTQSVVASEGSEVQMECYASGYPTPTITWRRENNAI 176

Qy     162 LDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSL 201
           |  : :      |     | |  |:| | ||| |:| : :
Db     177 LPTDSA------TYVGNTLPISSVEMEDPGTYYCVAVNGV 210


RESULT   6
Q9ERC8
ID   Q9ERC8      PRELIMINARY;      PRT;  2013 AA.
AC   Q9ERC8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DOWN SYNDROME CELL ADHESION MOLECULE.
GN   DSCAM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=BRAIN;
RA   Barlow G.M., Lyons G.E., Korenberg J.R.;
RT   "Down syndrome cell adhesion molecule, a highly conserved mouse
RT   homolog, dscam, is expressed in the neonatal and adult mouse brain.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR OUT-BRED;
RX   MEDLINE=21184302; PubMed=11237714;
RA   Agarwala K.L., Ganesh S., Amano K., Suzuki T., Yamakawa K.;
RT   "DSCAM, a highly conserved gene in mammals, expressed in
RT   differentiating mouse brain.";
RL   Biochem. Biophys. Res. Commun. 281:697-705(2001).
DR   EMBL; AF415558; AAG28796.1; -.
DR   EMBL; AY005483; AAF99440.1; -.
DR   MGD; MGI:1196281; Dscam.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 8.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00409; IG; 9
DR   SMART; SM00408; IGc2; 9.
DR   SMART; SM00410; IG_like; 6.
KW   Immunoglobulin domain.
SQ   SEQUENCE   2013 AA;  222268 MW;  1F4AF07947BEBA77 CRC64;


  Query Match             12.9%;  Score 172;  DB 11;  Length 2013;
  Best Local Similarity   29.8%;  Pred. No. 4.6e-06;
  Matches   57;  Conservative   33;  Mismatches   67;  Indels   34;  Gaps   10;

Qy      23 PREMISSVLIVNGKIENYTLDTIPGSQASLPCAVQNHTREEEELIWYREEGRVDLKSGNKI 82
           |||: |||                  ||| || |:|  :  ::|| |||  | : |  |  :
Db     329 PRKVESSV----------------GSQVSLSCSVIGN-EDVELSWYP-NGKI-INPGKNV 360

Qy      83 -----NSSSVCVSSISENDNGISFTCRLGRDQSVS---VSVVLNVTFPPLLSGNDFQTVE 134
                | ::: :  : ::| | :: | : :|:  :    | |||    | ::|    : |
Db     361 PIIGLNHANLIMDHMVKSDAG-AYLCFVPKDKLSAQDYVQVVLEDGIPKIISAFSEKVVS 419

Qy     135 FGSNVKLVCNVKANPQAQMMWYKNSSIIDFKSRHQIQQTFESE---FQGSLTAVEKPD 189
           | |||||| |  : | :  :  : |: | |:| |  |||      |:|: :  |
Db     420 PAEPVSIVCNVKSTPITPTVTWTIDGDPI-TKGSGHRISQMIISEGNVVSYLNISSSQVRD 478

Qy     190 NGIYSCIAKSS 200
           | | | | :|
Db     479 GGVYRCTANNS 489


RESULT   7
Q07153
ID   Q07153      PRELIMINARY;      PRT;  946 AA.
AC   Q07153;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RECEPTOR TYROSINE KINASE.
OS   Torpedo californica (Pacific electric ray).
OC   Eukaryota; Metazoa, Chordata, Craniata, Vertebrata, Chondrichthyes;
OC   Elasmobranchii; Squalea; Hyposqualea; Pristiorajea; Batoidea;
OC   Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
OX   NCBI_TaxID=7787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ELECTRIC ORGAN, AND MUSCLE;
RX   MEDLINE=93219391; PubMed=8385349;
RA   Jennings C.G.B., Dyer S.M., Burden S.J.;
RT   "Muscle-specific trk related receptor with a kringle domain defines a
RT   distinct class of receptor tyrosine kinases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:2895-2899(1993).
DR   EMBL; L11311; AAA49285.1; -.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000024; Fz_domain.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF01392; Fz; 1.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00051; kringle; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00018; KRINGLE.
DR   PRINTS; PR00109; TYRKINASE.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00130; KR; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50038; FZ; 1.
DR   PROSITE; PS00021; KRINGLE_1; FALSE_NEG.
DR   PROSITE; PS50070; KRINGLE_2; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Immunoglobulin domain; Kinase; Transferase;
KW   Tyrosine-protein kinase.
SQ   SEQUENCE   946 AA;  105892 MW;  81A9131DB7479D01 CRC64;


  Query Match             12.8%;  Score 171.5;  DB 13;  Length 946;
  Best Local Similarity   25.1%;  Pred. No. 2e-06;
  Matches   63;  Conservative   38;  Mismatches   95;  Indels   55;  Gaps   10;
```

```
Qy      15 LLLVIL·····FLPR··EMISSVLIVNGKTENYILDTTPGSQASLICAVQNHTREE 63
           ||:: |          ||:  ::|| : ||:   |          :|| :||| ::
Db      10 LLMIFLVTTGGSADGILPKAPQITSPLETVDALVE---------EEASFMCAVDSYP-AA 59

Qy      64 ELLWYREEGRV---DLKSGNKINSSSVCVSSISENDNGI-SFTCRLGRDQSVSVSVVLNV 119
           |: | |    :   | :   | |   : : |: : |||:   |   |   |      | |
Db      60 EITWTRNNIPIRPFDTRYSTKENGQILTILSVEDTDNGVYCCTANNGMGSSAQSCGALQV 119

Qy     120 TFPP--LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSES 177
             |  :   | : :  || | | |:   ||:  : |:|: : |  :: |  :    ||
Db     120 KMKPKIIRPPTDVRAL-LGSKVVLPCSTMGNPKPAISWFKDETALKNDQPRTSV---LES 175

Qy     178 FQLSITKVEKPDNGTYSCIAKSSLKTE----------------------SLDFHLIVK 213
             | |  |:  | | | |:|::||  |                        |    :|::
Db     176 GNLRIRNVQLEDAGKYRCLARNSLGFEYSRSAALEVQVSARIVKAPTSQNVSYGSEVILQ 235

Qy     214 DKTVGVPIEPI 224
           |  | ||  |
Db     236 CKATGFPIPTI 246


RESULT   8
Q9NYZ4
ID   Q9NYZ4      PRELIMINARY;      PRT;   431 AA.
AC   Q9NYZ4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIGLEC SAF2.
GN   SAF2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kikly K.K., Bochner B.S., Freeman S.D., Tan K.B., Gallagher K.T.,
RA   D'Alessio K.J., Holmes S.D., Abrahamson J.A., Erickson-Miller C.L.,
RA   Murdock P.R., Tachimoto H., Schleimer R.P., White J.R.;
RT   "Identification of SAF-2, a novel siglec expressed on eosinophils,
RT   mast cells and basophils.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF223403; AAF34702.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   431 AA;  46433 MW;  49B91B84CE67B118 CRC64;


  Query Match             12.7%;  Score 169.5;  DB 4;  Length 431;
  Best Local Similarity   23.6%;  Pred. No. 1.2e-06;
  Matches   63;  Conservative   53;  Mismatches   92;  Indels   59;  Gaps   11;

Qy      24 REMTSSVLTVNGKTENYILDT-TPGSQASLICAVQNHTREEELLWYREEGRVDLKS--GN 80
           ::::  |  :  : :  || |   |   :| |:|          |  ::|   : |  |
Db     144 KQLSVFVTALTHRPDILILGTLESGHSRNLTCSVP---------WACKQGTPPMISWIGA 194

Qy      81 KINS-------SSVCVSSISENDNGISFTCRL---GRDQSVSVSVVLNVTFPP------- 123
            ::|       |||   :    |:| | |::   |   : : :| |:|::||
Db     195 SVSSPGPTTARSSVLTLTPKPQDHGTSLTCQVTLPGTGVTTTSTVRLDVSYPPWNLTMTV 254

Qy     124 ---------LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQT 174
                    |      :| || ···||| | ·|| |·· | · |  | |  ||     ·
Db     255 FQGDATASTALGNGSSLSVLEGQSLRLVCAVNSNPPARLSWTRGS--LTLCPSR-----S 307

Qy     175 SESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPIIAACV--- 229
                |    | : :|   | | ::| |:::  :: :   | ::::  |   |:  :  | |
Db     308 SNPGLLELPRVHVRDEGEFICKAQNAQGSQHISLSLSLQNEGTGTSRPVSQVILAAVGGA 367

Qy     230 ----VIFLTLCFGLIARRKKIMKLCMK 252
                : |||: |   |     |:: | |
Db     368 GATALAFLSFCIIFI-----IVPSCPK 389


RESULT   9
Q921P2
ID   Q921P2      PRELIMINARY;      PRT;   605 AA.
AC   Q921P2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROTEIN FOR MGC:12077).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011310; AAH11310.1; -.
SQ   SEQUENCE   605 AA;  67353 MW;  DBD9BC64D245A867 CRC64;


  Query Match             12.4%;  Score 166;  DB 11;  Length 605;
  Best Local Similarity   28.0%;  Pred. No. 3.5e-06;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      49 QASLIC-------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 94
           | :| |               : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEKTL----DGHMVVRSHARV--SSLTLKSIQY 378

Qy      95 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 150
           | |  : |      :|:|   | |: | | : | | |        ||: | : | | | |
Db     379 TDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     151 AQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 209
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     210 LIVKDKTVGVP----IEPIIAACVVIF 232
           |:  | |   |    :||  :   | |
Db     495 LVQAD-TPSSPSIDRVEPYSSTAQVQF 520


RESULT  10
Q24273
ID   Q24273      PRELIMINARY;      PRT;  1011 AA.
AC   Q24273;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROMUSCULIN.
GN   NRM OR CG8779.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94000831; PubMed=8398154;
RA   Kania A., Han P.L., Kim Y.T., Bellen H.;
RT   "Neuromusculin, a Drosophila gene expressed in peripheral neuronal
RT   precursors and muscles, encodes a cell adhesion molecule.";
RL   Neuron 11:673-687(1993).
DR   EMBL; L23146; AAA03750.1; -.
DR   FlyBase; FBgn0005629; nrm.
DR   InterPro; IPR003599; Ig.
```

```
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 9.
DR   SMART; SM00409; IG; 4.
DR   SMART; SM00410; IG_like; 2.
KW   Cell adhesion.
SQ   SEQUENCE   1011 AA;  114125 MW;  AFDBA2A015D3AF63 CRC64;


  Query Match             12.1%;  Score 161;  DB 5;  Length 1011;
  Best Local Similarity   24.3%;  Pred. No. 1.8e-05;
  Matches   46;  Conservative   34;  Mismatches   81;  Indels   28;  Gaps    6;

Qy      31 LTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-----EEGRVDL--------- 76
           : :||   :    |  ||:  | | :        | ||        |  ::
Db     265 IDINGVKHH----TVQGSKVVLTCDIHGARPAVNLTWYNTTTIISSGENEITEVRSKSLE 320

Qy      77 KSGNKINSSSVCVSSISENDNGISFTC-------RLGRDQSVSVSVVLNVTFPPLLSGND 129
           ||    :: |  : : :  :|   | |       :: |:: :| :: | | :|||::  :
Db     321 KSDGIFHIQSELIFNATKEENDRVFFCEAENIVLQINREKPISSALTLEVLYPPVVKVSP 380

Qy     130 PQTVPEGSNVKLV-CNVKANPQA--QMMWYKNSSLIDLEKSRHQIQQTSESPQLSITKVE 186
           | : |: |   ||| :  |: ||:|  |::  : |    ||:  | |   |
Db     381 SAITANTSEIVLLNCEYFANPASLTQVEWYRNDILVNVNDTTHYKGGNSENVALVIKSTE 440

Qy     187 KPDNGTYSC 195
           | | | |||
Db     441 KEDIGNYSC 449


RESULT  11
Q9VNP2
ID   Q9VNP2      PRELIMINARY;      PRT;  1079 AA.
AC   Q9VNP2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NRM PROTEIN.
GN   NRM OR CG8779.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pteryqota; Neoptera; Endopteryqota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003599; AAF51887.1; -.
DR   FlyBase; FBgn0005629; nrm.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 9.
DR   SMART; SM00409; IG; 4.
DR   SMART; SM00410; IG_like; 2.
SQ   SEQUENCE   1079 AA;  121454 MW;  C7656582214C127B CRC64;


  Query Match             12.1%;  Score 161;  DB 5;  Length 1079;
  Best Local Similarity   24.3%;  Pred. No. 1.9e-05;
  Matches   46;  Conservative   34;  Mismatches   81;  Indels   28;  Gaps    6;

Qy      31 LTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-----EEGRVDL--------- 76
           : :||   :    |  ||:  | | :        | ||        |  ::
Db     265 IDINGVKHH----TVQGSKVVLTCDIHGARPAVNLTWYNTTTIISSGENEITEVRSKSLE 320

Qy      77 KSGNKINSSSVCVSSISENDNGISFTC-------RLGRDQSVSVSVVLNVTFPPLLSGND 129
           ||    :: |  : : :  :|   | |       :: |:: :| :: | | :|||::  :
Db     321 KSDGIFHIQSELIFNATKEENDRVFFCEAENIVLQINREKPISSALTLEVLYPPVVKVSP 380

Qy     130 PQTVPEGSNVKLV-CNVKANPQA--QMMWYKNSSLIDLEKSRHQIQQTSESPQLSITKVE 186
           | : |: |   ||| :  |: ||:|  |::  : |    ||:  | |   |
Db     381 SAITANTSEIVLLNCEYFANPASLTQVEWYRNDILVNVNDTTHYKGGNSENVALVIKSTE 440

Qy     187 KPDNGTYSC 195
           | | | |||
Db     441 KEDIGNYSC 449


RESULT  12
Q95LH0
ID   Q95LH0      PRELIMINARY;      PRT;  597 AA.
AC   Q95LH0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIALIC ACID-BINDING LECTIN SIGLEC-11.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21523976; PubMed=11546777;
RA   Angata T., Varki N.M., Varki A.;
RT   "A second uniquely human mutation affecting sialic acid biology.";
RL   J. Biol. Chem. 276:40282-40287(2001).
DR   EMBL; AF293372; AAL09302.1; -.
KW   Lectin.
SQ   SEQUENCE   597 AA;  65075 MW;  74C13CFA9CDB5BA5 CRC64;


  Query Match             11.9%;  Score 159;  DB 6;  Length 597;
  Best Local Similarity   23.2%;  Pred. No. 1.4e-05;
  Matches   63;  Conservative   45;  Mismatches   85;  Indels   78;  Gaps   12;

Qy      39 NYILD--------------TTPGSQAS-----LICAVQNHTREEELLWYREEGRVDLKS 78
```

```
          ||| |                : ||:  |     | |:|          |  |:|     :
Db    260 NYIYDKLSVHVTALTHLFTFSIPGTLESGHPKNLICSVF- -       WACEQGTPPTII 310

Qy     79 --GNKINSSSVCVSSIS-------ENDNGISFTCRL---GRDQSVSVSVVLNVTFPP--- 123
          |  ::|    :|  |          |:| | ||::   |   : : :| ||:::||
Db    311 WMGASVSSLEPTISRSSMLSLIPKPQDHGTSLTCQVTLPGAGVTTTRAVRLNISYPPQNL 370

Qy    124 ------------LLSGNDFQIVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSKHQ 170
                       |      :| || :: ||| | :|| |:: |   |  | |  |
Db    371 TMTVFQGDGTASTTLRNGSALSVLEGQSLHLVCAVDSNPPARLSWTWGS--LTLSPS--- 425

Qy    171 IQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPI---- 224
          |:|    | : :|   | | ::| |:: | :: :   | ::::  |   ||  :
Db    426 --QSSNLGVLELPRVHVKDEGEFTCRAQNPLGSQHISLSLSLQNEYTGKMRPISGVTLGA 483

Qy    225 ---IAACVVIFLTLCFGLIARRKKIMKLCMK 252
              |  ::||: |   :     ::: | |
Db    484 VGGAGATALVFLSFCIIFV-----VVRSCRK 509


RESULT  13
O55124
ID   O55124      PRELIMINARY;      PRT;  1463 AA.
AC   O55124;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   M-PROTEIN.
GN   MYOM2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98234547; PubMed=9570952;
RA   Steiner F., Weber K., Fuerst D.O.;
RT   "Structure and expression of the gene encoding murine M-protein, a
RT   sarcomere-specific member of the immunoglobulin superfamily.";
RL   Genomics 49:83-95(1998).
DR   EMBL; AJ001038; CAA04492.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:1328358; Myom2.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 4
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1463 AA;  164518 MW;  8C6551ED90F0B288 CRC64;


  Query Match             11.8%;  Score 158;  DB 11;  Length 1463;
  Best Local Similarity   25.1%;  Pred. No. 5.2e-05;
  Matches   60;  Conservative   36;  Mismatches   97;  Indels   46;  Gaps   10;

Qy     29 SVLTVNGKT-ENYIL--DTTPGSQAS-----------LICAVQNHTREEELLWYREEGR 73
          ||| | ||  |: ||      |: ||           | | ::  | | :: || :| :
Db   1211 SVLEVGGKVYEDMILAMSRVCGASASPLKVLCTPEGIRLQCFMKYFTEEMKVSWYHKEAK 1270

Qy     74 VDLKSGNKINSSS----VCVSSISENDNG-ISFTCRLGRD------------------- 108
          :     :|  |     : :  ·| | ·|   |·|
Db   1271 ISSSEHMRIGGSEEMAWLQICEPTEKDKGKYTFEIFDGKDSHQRSLDLSGQAFDEAYAEF 1330

Qy    109 QSVSVSVVLNVTFPPLLSG-NDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKS 167
          | :  :         :: |  |  |: ||  : | | |  ||  :::|:||    |:| |
Db   1331 QQLKAAAFAEKNRGKVIGGLPDVVTIMEGKTLNLTCTVFGNPDPEVVWFKNDK--DIELS 1388

Qy    168 KHQIQQTSESFQLSIT··KVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGVP IEP 223
          | : :  :|  :|:|    |   |:| ||   |:    | :|  : |      :| | |
Db   1389 FHPIVKMFQSKYVSLTIQGVTAFDSGKYSINVKNKYGGEKIDVTVSVYKHGEKIPDISP 1447


RESULT  14
Q94155
ID   Q94155      PRELIMINARY;      PRT;  1415 AA.
AC   Q94155;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNC-40 (T19B4.6 PROTEIN).
GN   UNC-40 OR T19B4.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL;
RA   Chan S.S.Y., Zheng H., Su M.W., Wilk R., Killeen M.T., Hedgecock E.M.,
RA   Culotti J.G.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1248-1415 FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [3]
RP   SEQUENCE OF 1248-1415 FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Gattung S.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1248-1415 FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U70618; AAB17088.1; -.
DR   EMBL; U80438; AAB37634.1; -.
DR   HSSP; P40189; 1BQU.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 1.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1415 AA;  154436 MW;  B2D90F76F9FFB6A9 CRC64;


  Query Match             11.8%;  Score 157.5;  DB 5;  Length 1415;
  Best Local Similarity   27.8%;  Pred. No. 5.5e-05;
  Matches   58;  Conservative   35;  Mismatches   83;  Indels   33;  Gaps   10;
```

```
Qy      24 REMTSSVLTVNGKT--ENYILDTTP-------GSQASLICAVQNHTREEELLWYREEGR- 73
           |   :: |||  :|     :  |||        | :  | | | :  |  :: | ::  :
Db     227 KSSQIAMLIVTLEIVSNRLVPIIPRIQVVEGSDEEIILPCLVASLIR-PQVRWLKISRQI 285

Qy      74  VELKSGNKINSSSVFVSSISKNDNGTSEICMI-GNLQSVSVSVVINVTIPPPLISCNDFQ 131
            ||      ::  ||: ||  |  | |: :|||    | |:  :| : |  || ::
Db     286 IVDGVRIRRVGVSSILVSRASIEDSG-YYLCRASNNLGSISDRAVSVEVPAPPRIITEPPTT 344

Qy     132 TVF-EGSNVKLVC-NVKANPQAQMMWYKN-----SSLLDLEKSRHQIQQTSESFQLSIT 184
           |  | ::|:| |    | |:|:: ||||      |    :| :|           | |
Db     345 KVAVETADVELECGTAAARPEAEVNWYKNGEAIQGSEYFVIEPNR----------LRIL 393

Qy     184 KVEKPDNGTYSCIAKSSLKTESLDFHLIV 212
           | : |   | |||:: : :|     |:|
Db     394 GVVRADQAIYQCIAENDVGSEQASAQLLV 422


RESULT  15
Q9VP08
ID   Q9VP08      PRELIMINARY;      PRT;   467 AA.
AC   Q9VP08;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG7166 PROTEIN.
GN   CG7166.
OS   Drosophila melanogaster (fruit fly)
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota, Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha,
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003595; AAF51754.1; -.
DR   FlyBase; FBgn0037107; CG7166.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003599; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   467 AA;  53023 MW;  5D0DCCF9CD3BBB7E CRC64;


  Query Match             11.8%;  Score 157;  DB 5;  Length 467;
  Best Local Similarity   26.5%;  Pred. No. 1.6e-05;
  Matches   48;  Conservative   37;  Mismatches   72;  Indels   24;  Gaps    8;

Qy      32 EVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKSG--NKINSSSVCV 89
           |:     |  :    ||  :| |    :     : |:::    |: ||  :  :|||: :
Db     136 TLRALPHNGQVIAKKGSIVTLECKASGNP-VPTIEWKK----DVFSGPTHLSDSSTLIL 190

Qy      90 SSISENDNGISFTCKL--GRDQSVSVSVVLNVTFPPLLSGNDFQTVE-------EGSNVK 140
             ::  :  | :: |    |    ||: : | :  || :      |||       || :|:
Db     191 ENVDRHHAG-TYQCSADNGVKDRVSMDIQLTILSPPEI------TVEKSWVHASEGYDVE 243

Qy     141 LVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSS 200
           ||| |  :  ::|:|||:|| |||    |  :    : : | |   :  | | |||:| ::
Db     244 LVCIVHGDVNSEMLWYQNGFLLD ATDRKSMYFRDDRYSLIIRNIQITDFGNYSCVADNA 302

Qy     201 L 201
           |
Db     303 L 303


Search completed: September 17, 2002, 13:23:23
Job time: 805 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 10:01:52 ; Search time 375.67 Seconds
(without alignments)
3605.945 Million cell updates/sec

Title: US-09-863-823-5
Perfect score: 789
Sequence: 1 atggcatggaagagcagtgt.....acagtgaaacagctctatga 789

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	558.4	70.8	1018	22	AAS28788	Human immunoglobul
	2	441	55.9	1958	22	AAH35000	Human colon cancer
c	3	284	36.0	19315	22	AAS28883	Human immunoglobul
	4	277.4	35.2	338	22	AAS37585	Novel human diagno
	5	250.8	31.8	361	22	AAS38121	Novel human diagno
	6	249	31.6	328	22	AAS38280	Novel human diagno
	7	247.2	31.3	359	22	AAF65373	Novel human polynu
	8	203.4	25.8	359	22	AAF65369	Novel human polynu
	9	41.6	5.3	2949	21	AAD00024	Degenerate DNA enc
	10	39.2	5.0	4092	20	AAV70230	Human receptor tyr
	11	39.2	5.0	4710	22	AAH98595	Human EST-derived
	12	39.2	5.0	4710	22	AAH98610	Human EST derived
	13	39.2	5.0	8513	22	AAD14203	MESVK/EGFP/IKESNCA
	14	38.6	4.9	2869	18	AAT90472	Rat muscle-specifi
	15	38.6	4.9	2869	18	AAT87074	Rat Dmk receptor c
	16	36.2	4.6	2675	19	AAV19076	Fibroblast growth
	17	36.2	4.6	2675	20	AAX85614	Bek-like newt FGFR
	18	36.2	4.6	2681	19	AAV19078	Keratinocyte growt
	19	36.2	4.6	2681	20	AAX85616	Newt keratinocyte
c	20	36.2	4.6	5115	22	AAK83483	Human immune/haema
	21	36	4.6	10132	24	ABL34081	Human immune syste
	22	35.8	4.5	903	14	AAQ47373	Partial sequence o
	23	35.8	4.5	903	17	AAT15762	Partial cDNA seque
	24	35.8	4.5	2526	22	AAF30917	Human follistatin-
	25	35.8	4.5	3503	19	AAV44041	Mouse bFGF recepto
	26	35.6	4.5	1382	23	AAS74298	DNA encoding novel
c	27	35.2	4.5	19653	24	ABL33334	Human immune syste
	28	35.2	4.5	235033	19	AAV57926	Hereditary haemoch
	29	35.2	4.5	237326	19	AAV57903	Hereditary haemoch
	30	35	4.4	789	21	AAC66560	Rat FGFR-1 coding
	31	35	4.4	915	21	AAC66561	Rat FGFR-1 flagged
	32	35	4.4	1983	22	AAD16027	Mouse soluble high
	33	35	4.4	2526	24	ABA01689	Mouse fibroblast g
	34	35	4.4	1503900	22	AAK95240	Human neuregulin-1
	35	35	4.4	1503900	22	AAK96733	Human neuregulin-1
	36	34	4.3	718	23	AAS77231	DNA encoding novel
	37	34	4.3	718	23	AAS77476	DNA encoding novel
	38	34	4.3	855	23	AAS55438	Streptococcus pneu
c	39	34	4.3	1628	21	AAC74342	Human secreted pro
c	40	33.4	4.2	193	21	AAC08794	Human secreted pro
	41	33.4	4.2	3734	23	AAS72406	DNA encoding novel
	42	33.4	4.2	3894	20	AAX55769	C. elegans Robo po
	43	33.4	4.2	3894	20	AAX57252	C. elegans ROBO cD
	44	33.4	4.2	5332	22	AAK52944	Human polynucleoti
	45	33.4	4.2	5489	22	AAK51960	Human polynucleoti

ALIGNMENTS

RESULT 1
AAS28788/c
ID AAS28788 standard; cDNA; 1018 BP.
XX
AC AAS28788;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immunoglobulin encoding cDNA SEQ ID No 34.
XX
KW Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;
KW antisense therapy, gene therapy, neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder, immune system disorder, proliferative disorder;
KW muscular disorder.
XX
OS Homo sapiens.
XX
PN WO200155315-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01326.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180528.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216847.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0240437.
PR 06-SEP-2000; 2000US-0240438.
PR 08-SEP-2000; 2000US-0241242.
PR 08-SEP-2000; 2000US-0241243.
PR 08-SEP-2000; 2000US-0241244.
PR 08-SEP-2000; 2000US-0241413.
PR 08-SEP-2000; 2000US-0241414.
PR 08-SEP-2000; 2000US-0242080.
PR 08-SEP-2000; 2000US-0242081.
PR 12-SEP-2000; 2000US-0241968.
PR 14-SEP-2000; 2000US-0242397.
PR 14-SEP-2000; 2000US-0242398.
PR 14-SEP-2000; 2000US-0242399.
PR 14-SEP-2000; 2000US-0242400.
PR 14-SEP-2000; 2000US-0242401.
PR 14-SEP-2000; 2000US-0243063.
PR 14-SEP-2000; 2000US-0243064.
PR 14-SEP-2000; 2000US-0243065.
PR 21-SEP-2000; 2000US-0244223.
PR 21-SEP-2000; 2000US-0244274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251039.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 05-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05 JAN 2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457725/49.
DR P-PSDB; AAU18000.
XX
PT Isolated novel immunoglobulin polypeptide for monitoring the presence
PT and progression of diseases and for diagnosis -
XX
PS Claim 1; SEQ ID No 34; 551pp; English.
XX
CC Sequences AAS28765-AAS28877 represent cDNA molecules, which encode the
CC immunoglobulin polypeptides of the invention, and primers for the
CC polynucleotides. The polynucleotides and polypeptides can be used to
CC diagnose a pathological condition or a susceptibility to a pathological
CC condition in a subject by determining the presence or absence of a

```
CC   mutation in a DNA sequence or determining the presence or amount of
CC   expression of the protein. Alternatively the identification of a binding
CC   partner to a sequence allows determination of changes in protein
CC   activity. The sequences can be used as research tools for receptors or
CC   other signal transduction pathway proteins that interact with the
CC   polypeptides of the invention and can be used to treat, prevent or
CC   diagnose various types of disorders such as neurological disorders,
CC   cardiovascular disorders, gastrointestinal disorders, reproductive
CC   disorders, immune system disorders, renal disorders, muscular disorders,
CC   pulmonary disorders, proliferative disorders and cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1018 BP; 284 A; 225 C; 198 G; 311 T; 0 other;


  Query Match             70.8%;  Score 558.4;  DB 22;  Length 1018;
  Best Local Similarity   99.1%;  Pred. No. 1.2e-164;
  Matches  559;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      79 acaagttctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacct 138
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     963 AGAGGTTCTGTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCT 904

Qy     139 ggctcccaagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaaactgctc 198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     903 GGCTCCCAAGCATCTCTGATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTC 844

Qy     199 tggtaccgagaggaggggagagtggatttgaaatctgsaaacaamatcaattccagctct 258
           ||||||||||||||||||||||||||||||||||||:||||||:||||||||||||||||
Db     843 TGGTACCGAGAGGAGGGGAGAGTGGATTTGAAATCTGGAAACAAAATCAATTCCAGCTCT 784

Qy     259 gtctgtgtctcttccatcagtgaaaatgacaacggaatcagctttacctgcaggctgggg 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     783 GTCTGTGTCTCTTCCATCAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGG 724

Qy     319 agggatcagtccgtgtccgtttcggtggtgctgaatgttacttttcctcctctcctaagt 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     723 AGGGATCAGTCCGTGTCCGTTTCGGTGGTGCTGAATGTTACTTTTCCTCCTCTCCTAAGT 664

Qy     379 ggaaacgacttccaaacagttgaggaaggcagtaatgtgaagttggtttgcaatgtgaaa 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     663 GGAAACGACTTCCAAACAGTTGAGGAAGGCAGTAATGTGAAGTTGGTTTGCAATGTGAAA 604

Qy     439 gccaacccccaggctcaaatgatgtggtacaaaaaacagtagtctcctcgatttagagaaa 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 GCCAACCCCCAGGCTCAAATGATGTGGTACAAAAACAGTAGTCTCCTCGATTTAGAGAAA 544

Qy     499 agccgtcaccaaatccaacagacaagtgagtcttttcagctgtcaatcaccaaagtcgag 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 AGCCGTCACCAAATCCAACAGACAAGTGAGTCTTTTCAGCTGTCAATCACCAAAGTCGAG 484

Qy     559 aagcctgacaacggaacctacagttgtattgcaaagtcatctctgaaaacggagagcttg 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 AAGCCTGACAACGGAACCTACAGTTGTATTGCAAAGTCATCTCTGAAAACGGAGAGCTTG 424

Qy     619 gactttcacctgattgttaaagat 642
           ||||||||||||||||||||||| |
Db     423 GACTTTCACCTGATTGTTAAAGCT 400


RESULT   2
AAH35000
ID   AAH35000 standard; cDNA; 1958 BP.
XX
AC   AAH35000;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:2082.
XX
KW   Human; colon cancer; colon cancer antigen, diagnosis, detection,
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG75595.
XX
PT   Nucleic acids encoding 4277 human colon cancer associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 3568-3569; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 999 to 982 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 1958 BP; 509 A; 480 C; 462 G; 495 T; 12 other;


  Query Match             55.9%;  Score 441;  DB 22;  Length 1958;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-127;
  Matches  441;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     349 ctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggc 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     644 ctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggc 703

Qy     409 agtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtac 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     704 agtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtac 763

Qy     469 aaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgag 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     764 aaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgag 823

Qy     529 tcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtatt 588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     824 tcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtatt 883

Qy     589 gcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagataaaact 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     884 gcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagataaaact 943

Qy     649 gtgggtgtaccaatagagcccattattgctgcatgtgttgtgatctttctgacattgtgc 708
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     944 gtgggtgtaccaatagagcccattattgctgcatgtgttgtgatctttctgacattgtgc 1003

Qy     709 tttgqactgattgctagaagaaaqaaaataatgaagctctgcatgaaggataaaqaccct 768
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1004 tttgqactgattgctagaagaaaqaaaataatgaagctctgcatgaaggataaaqaccct 1063

Qy     769 cacaqtqaaacaqctctatga 789
                ||||||||||||||||||||
Db    1064 cacaqtqaaacaqctctatga 1084


RESULT    3
AAS28883/c
ID   AAS28883 standard; DNA; 19315 BP.
XX
AC   AAS28883;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immunoglobulin encoding genomic DNA SEQ ID No 245
XX
KW   Immunoglobulin; signal transduction pathway protein; cancer; ds;
KW   antisense therapy; gene therapy; neurological disorder; renal disorder;
KW   cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW   reproductive disorder; immune system disorder; proliferative disorder;
KW   muscular disorder.
XX
OS   Homo sapiens.
XX
PN   WO200155315 A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01426
XX
PR   31-JAN-2000; 2000US-0179065
PR   04-FEB-2000; 2000US-0180628
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186450
PR   16-MAR-2000; 2000US-0189874
PR   17-MAR-2000; 2000US-0190076
PR   18-APR-2000; 2000US-0198123
PR   19-MAY-2000; 2000US-0205515
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215145.
PR   07-JUL-2000; 2000US-0216647
PR   07-JUL-2000; 2000US-0216880
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290
PR   26-JUL-2000; 2000US-0220963
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519
PR   14-AUG-2000; 2000US-0225213
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266
PR   14-AUG-2000; 2000US-0225267
PR   14-AUG-2000; 2000US-0225268
PR   14-AUG-2000; 2000US-0225278
PR   14-AUG-2000; 2000US-0225447
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758
PR   14-AUG-2000; 2000US-0225759
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681
PR   22-AUG-2000; 2000US-0226868
PR   22-AUG-2000; 2000US-0227182
PR   23-AUG-2000; 2000US-0227699
PR   30-AUG-2000; 2000US-0228924
PR   01-SEP-2000; 2000US-0229287
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   28-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
```

PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457725/49.
XX
PT Isolated novel immunoglobulin polypeptide for monitoring the presence
PT and progression of diseases and for diagnosis -
XX
PS Claim 1; SEQ ID No 245; 551pp; English.
XX
CC Sequences AAS28878-AAS28926 represent genomic DNA molecules which encode
CC the immunoglobulin polypeptides of the invention. The polynucleotides and
CC polypeptides can be used to diagnose a pathological condition or a
CC susceptibility to a pathological condition in a subject by determining
CC the presence or absence of a mutation in a DNA sequence or determining
CC the presence or amount of expression of the protein. Alternatively the
CC identification of a binding partner to a sequence allows determination of
CC changes in protein activity. The sequences can be used as research tools
CC for receptors or other signal transduction pathway proteins that interact
CC with the polypeptides of the invention and can be used to treat, prevent
CC or diagnose various types of disorders such as neurological disorders,
CC cardiovascular disorders, gastrointestinal disorders, reproductive
CC disorders, immune system disorders, renal disorders, muscular disorders,
CC pulmonary disorders, proliferative disorders and cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 19315 BP; 5674 A; 4041 C; 4034 G; 5565 T; 1 other;

Query Match 36.0%; Score 284; DB 22; Length 19315;
Best Local Similarity 91.2%; Pred. No. 6.9e-78;
Matches 299; Conservative 2; Mismatches 27; Indels 0; Gaps 0;

Qy 36 aagatttcttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaac 95
Db 14845 AGGTTGTTTACTCTGTATATCTAAATAGTTTCTCCCTATACTCCCAGGTTCTGTTTTAAC 14786

Qy 96 tgtgaatggtaaaactgagaactatatcctggatactacacctggctcccaagcatctct 155
Db 14785 TGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCTGGCTCCCAAGCATCTCT 14726

Qy 156 gatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggg 215

Db 14725 GATATGTGCTGTTCAAAACCACACCAGAGAAGAACTGCTCTGGTACCGAGAGGAGGG 14666

Qy 216 gagagtggatttgaaatctggaaacaaaatcaattccagctctgtctgtgtctcttccat 275
Db 14665 GAGAGTGGATTTGAAATCTGGAAACAAAATCAATTCCAGCTCTGTCTGTGTCTCTTCCAT 14606

Qy 276 cagtgaaaatgacaacggaatcagctttacctgcaggctgggaggggatcagtccgtgtc 335
Db 14605 CAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGAGGGATCAGTCCGTGTC 14546

Qy 336 cgtttcggtggtgctgaatgttacttgt 363
Db 14545 CGTTTCGGTGGTGCTGAATGTTACTTGT 14518

RESULT 4
AAS37585
ID AAS37585 standard; cDNA; 338 BP.
XX
AC AAS37585;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #643.
XX
KW Human, cancer, breast, lung, colon, prostate, cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 757; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 338 BP; 93 A; 69 C; 91 G; 85 T; 0 other;

Query Match 35.2%; Score 277.4; DB 22; Length 338;
Best Local Similarity 98.2%; Pred. No. 9.8e-77;
Matches 278; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 79 acaagttctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacct 138
Db 56 agaggttctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacct 115

```
Qy     139 qqctcccaaqcatctctqatatqtqctqttcaaaaccacaccaqaqaggaaqaactqctc 198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 qqctcccaaqcatctctqatatqtqctqttcaaaaccacaccaqaqaggaaqaactqctc 175

Qy     199 tqqtaccqaqaqqaqqqqaqaqtqqatttqaaaatctqqaaacaaaatcaattccaqctct 258
           ||||||||||||||||||||||||||||||||||:||||||:||||||||||||||||||
Db     176 tqqtaccqaqaqqaqqqqaqaqtqqatttqaaatctqqaaacaaaatcaattccaqctct 235

Qy     259 qtctqtqtctcttccatcaqtqaaaatqacaacqqaatcaqctttacctqcaqqctqqqq 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 qtctqtqtctcttccatcaqtqaaaatqacaacqqaatcaqctttacctqcaqqctqqqq 295

Qy     319 aqqqatcaqtccqtqtccqtttcqqtqqtqctqaatqttactt 361
           ||||||||||||||||||||||| ||||||||||||||||||
Db     296 aqqqatcaqtccqtqtccqtttcqqaaqtqctqaatqttactt 338
```

RESULT 5
AAS38121
ID AAS38121 standard; cDNA; 361 BP.
XX
AC AAS38121;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #1179.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001 530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 892; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 361 BP; 114 A; 63 C; 90 G; 94 T; 0 other;

Query Match 31.8%; Score 250.8; DB 22; Length 361;
Best Local Similarity 98.8%; Pred. No. 2.2e-68;
Matches 249; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atggcatggaagagcagtgtcataatgcaaatgggaagatttcttctcttagtaatttta 60
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
```

```
Db     109 atggcatggaagagcagtgtcataatgcaaatgggaagatttcttctcttagtaatttta 168

Qy      61 tttctgccacgtgaqatgacaaqttctgttttaactgtgaatggtaaaactgaqaactat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     169 tttctgccacgtgaqatgacaaqttctgttttaactgtgaatggtaaaactgaqaactat 228

Qy     121 atcctqqatactacaccqqctcccaaqcatctctqatatqtqctqttcaaaaccacacc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     229 atcctqqatactacaccqqctcccaaqcatctctqatatqtqctqttcaaaaccacacc 288

Qy     181 aqaqaqqaaqaactqctctqqtaccqaqaqqaqqqqaqaqtqqatttqaaaatctqqaaac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db     289 aqaqaqqaaqaactqctctqqtaccqaqaqqaqqqqaqaqtqqatttqaaaatctqqaaac 348

Qy     241 aaatccaattcc 252
           ||:||||||||
Db     349 aaaatcaattcc 360
```

RESULT 6
AAS38280
ID AAS38280 standard; cDNA; 328 BP.
XX
AC AAS38280;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #1338.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001 530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 932; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 328 BP; 102 A; 62 C; 81 G; 83 T; 0 other;

Query Match 31.6%; Score 249; DB 22; Length 328;
Best Local Similarity 97.6%; Pred. No. 7.8e-68;
Matches 249; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 atggcatggaagagcagtgtcataatgcaratgggaagatttcttctcttagtaatttta 60
          ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db     73 atggcatggaagagcagtgtcataatgcaaatgggaagatttcttctcttagtaatttta 132

Qy     61 tttctgccacgtgagatgacaagttctgttttaactgtgaatggtaaaactgagaactat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    133 tttctgccacgtgagatgacaagttctgttttaactgtgaatggtaaaactgagaactat 192

Qy    121 atcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacacc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 atcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacacc 252

Qy    181 agagaggaagaactgctctggtaccgagaggaggggagagtggatttgaaatctgsaaac 240
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||:||||
Db    253 agagaggaagaactgctctggtaccgacaggaggggagagtggatttgaaatctggaaac 312

Qy    241 aamatcaattccagc 255
          ||:||| ||||| ||
Db    313 aaaatccattcccgc 327


RESULT   7
AAF65373
ID   AAF65373 standard; cDNA; 359 BP.
XX
AC   AAF65373;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1129.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 706; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 359 BP; 113 A; 64 C; 92 G; 89 T; 1 other;


  Query Match             31.3%;  Score 247.2;  DB 22;  Length 359;
  Best Local Similarity   98.4%;  Pred. No. 3e-67;
  Matches  246;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 atggcatggaagagcagtgtcataatgcaratgggaagatttcttctcttagtaatttta 60
          ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db    109 atggcatggaagagcagtgtcataatgcaaatgggaagatttcttctcttagtaatttta 168

Qy     61 tttctgccacgtgagatgacaagttctgttttaactgtgaatggtaaaactgagaactat 120
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    169 tttctgccacgtgagatgacaagttctgttttaactgtgaatcgtaaaactgagaactat 228

Qy    121 atcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacacc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    229 atcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacacc 288

Qy    181 agagaggaagaactgctctggtaccgagaggaggggagagtggatttgaaatctgsaaac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db    289 agagaggaagaactgctctggtaccgagaggaggggagagtggatttgaaatctggaaac 348

Qy    241 aamatcaatt 250
          ||:||||||
Db    349 aaaatcaatt 358


RESULT   8
AAF65369
ID   AAF65369 standard; cDNA; 359 BP.
XX
AC   AAF65369;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1125.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 705; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
```

```
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 359 BP; 106 A; 69 C; 95 G; 88 T; 1 other;


  Query Match             25.8%;  Score 203.4;  DB 22;  Length 359;
  Best Local Similarity   94.6%;  Pred. No. 1.7e-53;
  Matches  229;  Conservative    2;  Mismatches    9;  Indels    2;  Caps    2;

Qy       1 atqqcatqqaaaqaqcaqtqtcataatqcaratqqqaaaqatttcttctcttaqtaatttta 60
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||| ||
Db     109 atqqcatqqaaaqaqcaqtqtcataatqcaaatqqqaaaqatttcttctcttaqtaatctta 168

Qy      61 tttctqccacqtqaqatqacaacqttctqttttaactqtqaatqqtaaaactqaqaactat 120
           |||||||||||||||||||||||||||| |||||||||||| |||| ||||||||||||
Db     169 tttctqccacqtqaqatqacaaqttctqqttttaactqtqaatcqtaactqaqaactat 228

Qy     121 atcctqqatactacaccctqqctcccaaqcatctctqatatqtqctgttccaaaaccacacc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     229 atcctqqatactacaccctqqctcccaaqcatctctqatatqtqctgttccaaaaccacacc 288

Qy     181 aqaqaqqaaqaactqctctqqtaccqaqaqqaqqqqaqaqtqqatttqaaatctqsaaac 240
           |||||||||| |||||||||||||||||||| ||||||| |||||||| |||||:||
Db     289 aqaqaqqaaq actqctctqqtaccqaqaqqa qqqaqaqqqqatttqaaatctqqaaca 346

Qy     241 aa 242
           ||
Db     347 aa 348


RESULT   9
AAD00024
ID   AAD00024 standard; DNA; 2949 BP.
XX
AC   AAD00024;
XX
DT   24-JUL-2000  (first entry)
XX
DE   Degenerate DNA encoding follistatin-related protein zfsta.
XX
KW   Follistatin-related protein; zfsta2; Transforming growth factor-beta;
KW   TGF-beta family; acute inflammatory condition; bursitis; myositis;
KW   chronic inflammatory demyelinating polyneuropathy; contact dermatitis;
KW   contact vulvovaginitis; sepsis; ulcerative colitis; dermatological;
KW   antiinflammatory, antiulcer, antibacterial; immunosuppressive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200022126 A1.
XX
PD   20-APR-2000.
XX
PF   05-OCT-1999;   99WO-US23179.
XX
PR   15-OCT-1998;   98US-0173044
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC,  Ellsworth JL;
XX
DR   WPI; 2000-317983/27.
XX
PT   New polypeptide comprising follistatin homology domain is useful for
PT   treating acute inflammatory conditions, bursitis, contact dermatitis,
PT   contact vulvovaginitis, myositis, sepsis and ulcerative colitis
XX
PS   Disclosure, Page 117-118; 125pp; English.
XX
CC   The present sequence encodes follistatin related protein zfsta2. This is
CC   a degenerate DNA that encompasses all DNAs that encode zfsta2 protein.
CC   zfsta2 gene and its protein are useful for binding to members of the
CC   transforming growth factor (TGF)-beta family and mediating central
CC   nervous system, reproductive, haematopoietic and bone related activities.
CC   This is useful for treating acute inflammatory conditions, bursitis,
CC   chronic inflammatory demyelinating polyneuropathy, various forms of
CC   contact dermatitis, contact vulvovaginitis, myositis, sepsis and
CC   ulcerative colitis. zfsta2 cDNA is also useful as a primer and probe for
CC   identifying genes encoding zfsta2-like proteins.
XX
SQ   Sequence 2949 BP; 614 A; 274 C; 477 G; 396 T; 1188 other;


  Query Match              5.3%;  Score 41.6;  DB 21;  Length 2949;
  Best Local Similarity   33.4%;  Pred. No. 0.03;
  Matches  133;  Conservative   46;  Mismatches  153;  Indels    6;  Caps    2;

Qy     299 acqqaatcaqctttaactqcaqqctqqqqaqqqatcaqtccqtqtccqtttcqqtqqtqc 349
           |:|     | :| :|| ||:       |  |  |: |:   || | :    |    |
Db    1352 ayqtnqqnaaytayacntqytayqcnqayqqntayqarcarqtntayearacncayatht 1411

Qy     350 tqaatqttactttttcctcctctcctaaqtqqaaac  qacttccaaaacaqttqaqqaaq 406
           |  | || | : | || ||  |  |::| |   |:  || :  ||: |    ||:   |
Db    1412 tycarqtnaayqttncchccntqtnathmgngtttayccngarwsncarqcnmqnqarccnq 1471

Qy     407 gcagtaatgtgaaqttggtttgcaatqtqaaaqccaaccccccaggctcaaatqatqtggt 466
           |    |  |  :  :|    ||: |:|   |:|  | :||  |: | ||: |    |||:
Db    1472 gngtnacngctwsaytnmgntqycaygcnqarqgnathccnaarccncarytnqgntqgy 1531

Qy     467 acaaaaacaqtagtctcctcqatttagaqaaaaaqccqtcaccaaatccaacaqacaaqtg 526
             ||:||: | | :  : |:       |:   ::    || :  |  :  ||: | | :|
Db    1532 tnaaraayggnathgayathacnccnaarytnwsnaarcarytnacnytncarqcnaayg 1591

Qy     527 aqtctttttcaqctqtcaatcaccaaaqtcqagaaqcctqacaacqqaacctacaqttqta 586
             ::     |: |    ||::: || ||     |    ||:|  ||  | ||:|  ||:|
Db    1592 gtwstq   arqtncaytathwsnaaygtnmgntaygarqaycanqqnqcntayacntqya 1648

Qy     587 ttqcaaaq 594
           |:|| ||:
Db    1649 thqcnaar 1656


RESULT  10
AAV70230
ID   AAV70230 standard; DNA; 4092 BP.
XX
AC   AAV70230;
XX
DT   11-FEB-1999  (first entry)
XX
DE   Human receptor tyrosine kinase ROR-2 encoding DNA.
XX
KW   Receptor tyrosine kinase; Ror-1; Ror 2; Ehk 1; Ehk 2; detection;
KW   neurotrophin activity; trkB; proto-oncogene; tyrosine kinase receptor;
KW   binding protein; BDNF; NT-3; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             200..3031
FT                   /*tag= a
XX
PN   US5843749-A.
XX
```

```
PD   01-DEC-1998.
XX
PF   06-JUN-1995;   95US-0469537
XX
PR   17-MAR-1995;   95US-0406247
PR   26-JUL-1991;   91US-0736559
PR   28-OCT-1993;   93US-0144992
PR   06-JUN-1995;   95US-0469537.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Maisonpierre PC,  Masiakowski P,  Yancopoulos GD;
XX
DR   WPI; 1999-044584/04.
DR   P-PSDB; AAW83168.
XX
PT   DNA encoding receptor tyrosine kinase proteins - and corresponding
PT   proteins
XX
PS   Example; Fig 27; 194pp; English.
XX
CC   The present invention describes nucleic acid molecules for ror-1,
CC   ror-2, ehk-1 and ehk-2. Also described are the corresponding proteins:
CC   Ror-1; Ror-2; Ehk-1; and Ehk-2. The proteins are orphan receptor
CC   tyrosine kinases. The present sequence encodes human ROR-2.
XX
SQ   Sequence 4092 BP; 884 A; 1229 C; 1171 G; 808 T; 0 other;


  Query Match             5.0%;  Score 39.2;  DB 20;  Length 4092;
  Best Local Similarity   46.9%;  Pred. No. 0.2;
  Matches  122;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy     349 ctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggc 408
           ||||| | |   ||||    | | ||      |    ||      || |||   ||||| 
Db     368 ctgaaaggttactttctgaattttctggagccagtaaacaatatcaccattgtccaaggc 427

Qy     409 agtaatgtgaagttggtttgcaatgtgaaagccaacccccaggctcaaatgatgtggtac 468
             |  |  |   ||   ||||| |||  ||  ||||| |   || |  ||  ||||  
Db     428 cagacggcaattctgcactgcaaggtggcaggaaacccacccccctaacgtgcggtggcta 487

Qy     469 aaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgag 528
           || ||    |   |   | |     |||    | ||    |   ||||   |||||    | 
Db     488 aagaatgatgccccggtggtgcaggagccgcggcggatcatcatccggaagacagaatat 547

Qy     529 tcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtatt 588
             ||  |  |||  ||||    |  | || | | | |||| ||   ||||  || |  | 
Db     548 ggttcacgactgcgaatccaggacctggacacgacagacactggctactaccagtgcgtg 607

Qy     589 gcaaagtcatctctgaaaac 608
           || |        |||| || 
Db     608 gccaccaacgggatgaagac 627


RESULT  11
AAH98595
ID   AAH98595 standard; cDNA; 4710 BP.
XX
AC   AAH98595;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Human EST-derived coding sequence SEQ ID NO. 452.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02687.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   17-JUL-2000; 2000US-0617746.
PR   03-AUG-2000; 2000US-0631451.
PR   15-SEP-2000; 2000US-0663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Qian XB,  Wang Z,  Chen R,  Asundi V;
PI   Cao Y,  Drmanac RA,  Zhang J,  Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM23936.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Claim 1; Page 500-501; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention.
XX
SQ   Sequence 4710 BP; 1179 A; 1258 C; 1167 G; 1106 T; 0 other;


  Query Match             5.0%;  Score 39.2;  DB 22;  Length 4710;
  Best Local Similarity   51.1%;  Pred. No. 0.21;
  Matches   92;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy     415 gtgaagttggtttgcaatgtgaaagccaacccccaggctcaaatgatgtggtacaaaaac 474
           || |   |||| ||| |||   ||| |  |||  || |    ||||  |||   ||  | 
Db     836 gtcaccctggtgtgcgatgccgaaggcttcccagagcccaccatgagctggacaaaggat 895

Qy     475 agtagtctcctcgatttagaagaaaagccgtcaccaaatccaacagacaagtgagtctttt 534
             |    |   | ||   |||| ||  |  | |  |   |       |    ||   ||  
Db     896 ggggaacagatagagcaagaggaagacgatgagaagtacatcttcagcgacgataqttcc 955

Qy     535 cagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaag 594
           |||||| | ||||  || || || |||   ||| | |     ||||  || |||||  || 
Db     956 cagctgaccatcaaaaaggtggataagaacgacgaggctgagtacatctgcattgctgaq 1015


RESULT  12
AAH98610
ID   AAH98610 standard; cDNA; 4710 BP.
XX
AC   AAH98610;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Human EST-derived coding sequence SEQ ID NO: 467.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02687.
```

XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 04-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V,
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR P-PSDB; AAM23951.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 509-510; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SQ Sequence 4710 BP; 1174 A; 1258 C; 1167 G; 1106 T; 9 other;

Query Match 5.0%; Score 39.2; DB 22; Length 4710;
Best Local Similarity 51.1%; Pred. No. 0.21;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 415 gtqaaqttqgtttqcaatgtgaaaqccaaccccqaqqctcaaatgatgtggtacaaaaac 474
Db 836 gtcaccctggtgtgcgatgccgaaggcttcccagagcccaccatgagctggacaaaggat 895

Qy 475 agtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtctttt 534
Db 896 gggaaacagataqagcaagaggaagacgatgagaagtacatcttcagcgacgatagttcc 955

Qy 535 cagctgtcaatcaccaaaqtcgaqaaqcctgacaacggaacctacagttgtattgcaaag 594
Db 956 cagctgaccatcaaaaagtgtggataagaacgacgaggctgagtacatctgcattgctgag 1015

RESULT 14
AAD14203
ID AAD14203 standard; DNA; 8513 BP.
XX
AC AAD14203;
XX
DT 06-NOV-2001 (first entry)
XX
DE MESVR/EGFP/IRESNCAMPro(ori) vector.
XX
KW Transcriptional regulatory element; translational regulatory element;
KW gene expression, diagnostic application, therapeutic application;
KW MESVR; retrovirus vector; enhanced green fluorescent protein; EGFP; IRES;
KW internal ribosome entry sequence, N CAM, neural cell adhesion molecule,
KW ori; origin of replication; SV40; human; mouse; ds.
XX
OS Chimeric - Murine embryonic stem cell virus.
OS Chimeric - Rous sarcoma virus.
OS Chimeric - Mastadenovirus.
OS Chimeric - Mus sp.
OS Chimeric - Rhesus macaque polyoma virus.
OS Chimeric - Homo sapiens.
OS Chimeric - Synthetic.
XX
PN WO200155371-A1.

XX
PD 02-AUG-2001.
XX
PF 26-JAN-2001; 2001WO-US02733.
XX
PR 28-JAN-2000; 2000US-0178816.
PR 02-MAR-2000; 2000US-0186496.
PR 30-MAY-2000; 2000US-0207804.
PR 07-SEP-2000; 2000US-0230852.
PR 07-SEP-2000; 2000US-0230956.
PR 12-JAN-2001; 2001US-0261312.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Madre VF, Edelman GM, Chappell GM, Jones FS, Owens G, Meech R;
XX
DR WPI; 2001-483242/52.
XX
PT Identifying oligonucleotides with transcriptional/translational
PT regulatory activity in eukaryotic cells by integrating an
PT oligonucleotide into cell genome and detecting a change in expression
PT of expressible polynucleotides -
XX
PS Claim 95; Page 131-135; 172pp; English.
XX
CC The present invention relates to a method for the identification of an
CC oligonucleotide with transcriptional/translational regulatory activity in
CC a eukaryotic cell. The method involves integrating an oligonucleotide
CC into the genome of a eukaryotic cell so that it is linked to an
CC expressible polynucleotide or contacting the eukaryotic cell with library
CC of vectors, obtained by cloning a library of oligonucleotide into
CC multiple copies of expression vectors comprising polynucleotide, and
CC detecting changes in expression of the polynucleotide. The method is
CC useful for identifying an oligonucleotide having transcriptional or
CC translational activity in a eukaryotic cell. It is also useful for
CC identifying synthetic transcriptional or translational regulatory
CC elements. The transcriptional or translational regulatory elements are
CC useful in a variety of gene expression configurations for regulating
CC control of expression, and in expression vectors for controlling gene
CC expressions in diagnostic and therapeutic applications. The present
CC sequence is MESVR/EGFP/IRESNCAMPro(ori) vector based on
CC Murine embryonic stem cell virus (MESV) retrovirus. The vector comprises
CC nucleotide sequences encoding enhanced green fluorescent protein
CC (EGFP) and human neural cell adhesion molecule (N CAM) linked by an
CC internal ribosome entry sequence (IRES). The upstream long terminal
CC repeat (LTR) region of the retrovirus was modified to contain
CC Rous sarcoma virus (RSV) enhancer elements. The downstream LTR U3 region
CC was modified to contain a cassette containing a polylinker for the
CC insertion of random oligonucleotides, adenovirus major late promoter,
CC the initiator sequence (Inr) from mouse terminal deoxynucleotidyl
CC transferase gene and a complete R region. The vector also contains a
CC simian virus 40 (SV40) origin of replication (ori) and is useful
CC for identifying synthetic transcriptional regulatory elements.
XX
SQ Sequence 8513 BP; 2067 A; 2353 C; 2224 G; 1869 T; 9 other;

Query Match 5.0%; Score 39.2; DB 22; Length 8513;
Best Local Similarity 51.1%; Pred. No. 0.29;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 415 gtqaaqttqgtttqcaatgtgaaaqccaaccccqaqqctcaaatgatgtggtacaaaaac 474
Db 3607 gtcaccctggtgtgcgatgccgaaggcttcccagagcccaccatgagctggacaaaggat 3666

Qy 475 agtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtctttt 534
Db 3667 gggaaacagataqagcaagaggaagacgatgagaagtacatcttcagcgacgatagttcc 3726

Qy 535 cagctgtcaatcaccaaaqtcgaqaaqcctgacaacggaacctacagttgtattgcaaag 594
Db 3727 cagctgaccatcaaaaagtgtggataagaacgacgaggctgagtacatctgcattgctgag 3786

```
RESULT  14
AAT90472
ID   AAT90472 standard; cDNA; 2869 BP.
XX
AC   AAT90472;
XX
DT   27-JAN-1998  (first entry)
XX
DE   Rat muscle-specific kinase (MuSK) cDNA.
XX
KW   Receptor tyrosine kinase; muscle specific kinase; MuSK; Dmk; rat;
KW   ligand; agrin; diagnosis; therapy; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             135..2741
FT                   /*tag= a
FT   sig_peptide     135..191
FT                   /*tag= b
FT   mat_peptide     192..2738
FT                   /*tag= c
XX
PN   WO9721811-A2.
XX
PD   19-JUN-1997.
XX
PF   13-DEC-1996;   96WO-US20696.
XX
PR   10-MAY-1996;   96US-0644271.
PR   15-DEC-1995;   95US-0008657.
XX
PA   (REGE-) REGENERON PHARM INC
XX
PI   Bowen DC,  Glass DJ,  Valenzuela DM,  Yancopoulos GD;
XX
DR   WPI; 1997-332783/30.
DR   P-PSDB; AAW26610.
XX
PT   Nucleotide sequences encoding human agrin and muscle specific kinase
PT   and related receptor - used in diagnosis and treatment of disorder
PT   with muscle atrophy
XX
PS   Example 1; Fig 1; 120pp; English.
XX
CC   This cDNA sequence, deposited as ATCC 75498, includes a coding
CC   sequence for a novel rat receptor tyrosine kinase designated muscle
CC   specific kinase or MuSK (see AAW26610) that is expressed in normal
CC   and denervated muscle.  MuSK is alternatively referred to Dmk for
CC   denervated muscle kinase.  The cDNA clone was isolated from
CC   denervated muscle cDNA using primers (see AAT90476 and AAT90480)
CC   based on tyrosine kinase homology domains.  Human MuSK (see
CC   AAT90473) has also been identified.  Use of MuSK to generate anti-
CC   MuSK antibodies and in the diagnosis of neurological or other
CC   disorders is disclosed.  Assay systems that may be used to detect
CC   and/or measure ligands that bind the musk gene product are provided.
CC   A claimed method of promoting the growth, differentiation or
CC   survival of MuSK receptor-expressing cells involves administration
CC   to the cell of agrin (see AAW26609).  Such cells include muscle,
CC   heart, spleen, ovary and retina cells, or cells genetically
CC   engineered to express the MuSK receptor.
XX
SQ   Sequence 2869 BP; 727 A; 754 C; 726 G; 658 T; 4 other;


  Query Match             4.9%;  Score 38.6;  DB 18;  Length 2869;
  Best Local Similarity   55.3%;  Pred. No. 0.25;
  Matches   99;  Conservative    0;  Mismatches   74;  Indels    6;  Gaps    1;

Qy    415 gtgaagttggtttgcaatgtgaaagccaacccccaggctcaaatgatgtggtacaaaaac 474
          | ||  ||  | |||   ||| || || |||| ||| || ||||    |||  || |||
Db    267 gcgactttcatgtgcgccgtggaatcctaccctcagcctgaaatttcttggaccagaaat 326

Qy    475 agtagtctcctcgatttagaggaaaagccgtcaccaaaatccaacagacaagtgagtctttt 534
          |  | |||| || |  |      | | |||  ||   |||| | |||   || ||
Db    327 aaaattctcatcaagctgtttgacacccgctacagcatccgagagaacggtcag------ 380

Qy    535 cagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaa 593
          |  ||  | |||   |  || ||| |   |||  | || | |||| | || |  || ||
Db    381 ctcctcaccatcctgagtgtggaggacagtgatgatggcatctactgctgcacagccaa 439


RESULT  15
AAT87074
ID   AAT87074 standard; cDNA; 2869 BP.
XX
AC   AAT87074;
XX
DT   06-JAN-1998  (first entry)
XX
DE   Rat Dmk receptor cDNA.
XX
KW   Dmk receptor; tyrosine kinase receptor; signal transduction; assay;
KW   therapy; diagnosis; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             135..2741
FT                   /*tag= a
FT   sig_peptide     135..191
FT                   /*tag= b
FT   mat_peptide     192..2738
FT                   /*tag= c
XX
PN   US5656473-A.
XX
PD   12-AUG-1997.
XX
PF   21-JUL-1993;   93US-0095658.
XX
PR   19-JAN-1995;   95US-0374834.
PR   21-JUL-1993;   93US-0095658.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Rojas EA,  Valenzuela DM;
XX
DR   WPI; 1997-414593/38.
DR   P-PSDB; AAW26507.
XX
PT   New isolated human denervated muscle kinase receptor - used to
PT   develop products for the diagnosis and treatment of neurological,
PT   muscle or neuromuscular disorders
XX
PS   Example 1; Column 23-28; 31pp; English.
XX
CC   This cDNA clone, deposited as ATCC 75498,codes for rat Dmk receptor
CC   (AAW26507), a novel tyrosine kinase receptor that is expressed at high
CC   levels in denervated muscle.  It was isolated from a rat denervated
CC   skeletal muscle cDNA library following isolation of a partial
CC   sequence using primers (see AAT87075 and AAT87081) based on receptor
CC   tyrosine kinase (RTK) homology domains.  PCR primers (AAT87075-81)
CC   based on homology domains between novel Dmk and RTKs can be used to
CC   identify novel RTKs.  A human clone (see AAT87073) coding for a
CC   claimed human Dmk receptor (AAW26506) has been obtained.
XX
SQ   Sequence 2869 BP; 727 A; 754 C; 726 G; 658 T; 4 other;


  Query Match             4.9%;  Score 38.6;  DB 18;  Length 2869;
  Best Local Similarity   55.3%;  Pred. No. 0.25;
  Matches   99;  Conservative    0;  Mismatches   74;  Indels    6;  Gaps    1;

Qy    415 gtgaagttggtttgcaatgtgaaagccaacccccaggctcaaatgatgtggtacaaaaac 474
```

```
Db     267 gcqactttcatgtgcgccgtgggatcctacctcaqcctgaaatttcttggaccagaaat 326

Qy     475 agtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtctttt 534
Db     327 aaaattctcatcaagctgtttgacaccgctacagcatccgagagaacggtcag------ 380

Qy     535 cagctgtcaatcaccaaagtcgagaagcctgaccagtggaacctacagttgtattgcaaa 593
Db     381 ctcctcaccatcctgagtgtggaggacagtgatgatggcatctactgctgcacagccaa 439
```

Search completed: September 17, 2002, 10:02:08
Job time: 6036 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 08:21:32 ; Search time 375.67 Seconds
(without alignments)
3496.259 Million cell updates/sec

Title: US-09-863-823-1
Perfect score: 765
Sequence: 1 atgcaaatgggaagatttct.....acagtgaaacagctctatga 765

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 1472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	558.4	73.0	1018	22	AAS28788	Human immunoglobul
	2	441	57.6	1958	22	AAH35000	Human colon cancer
c	3	284	37.1	19315	22	AAS28883	Human immunoglobul
	4	277.4	36.3	338	22	AAS37585	Novel human diagno
	5	227.2	29.7	361	22	AAS38121	Novel human diagno
	6	225.4	29.5	328	22	AAS38280	Novel human diagno
	7	223.6	29.2	359	22	AAP65373	Novel human polynu
	8	179.8	23.5	359	22	AAP65369	Novel human polynu
	9	41.6	5.4	2949	21	AAD00024	Degenerate DNA enc
	10	39.2	5.1	4092	20	AAV70230	Human receptor tyr
	11	39.2	5.1	4710	22	AAH98595	Human EST-derived
	12	39.2	5.1	4710	22	AAH98610	Human EST-derived
	13	39.2	5.1	8513	22	AAD14203	MFSVR/FGFP/IRESNCA
	14	38.6	5.0	2869	18	AAT90472	Rat muscle-specifi
	15	38.6	5.0	2869	18	AAT87074	Rat Dmk receptor c
	16	36.2	4.7	2675	19	AAV19076	Fibroblast growth
	17	36.2	4.7	2675	20	AAX85614	Bek-like newt FGFR
	18	36.2	4.7	2681	19	AAV19078	Keratinocyte growt
	19	36.2	4.7	2681	20	AAX85616	Newt keratinocyte
c	20	36.2	4.7	5115	22	AAK83483	Human immune/haema
	21	35.8	4.7	903	14	AAQ47379	Partial sequence o
	22	35.8	4.7	903	17	AAT15762	Partial cDNA seque
	23	35.8	4.7	2526	22	AAF30917	Human follistatin-
	24	35.8	4.7	3503	19	AAV44041	Mouse bFGF recepto
	25	35.6	4.7	1382	23	AAS74298	DNA encoding novel
	26	35.6	4.7	235033	19	AAV57926	Hereditary haemoch
	27	35.6	4.7	237326	19	AAV57903	Hereditary haemoch
c	28	35.2	4.6	19653	24	ABL33334	Human immune syste
	29	35	4.6	789	21	AAC66560	Rat FGFR-1 coding
	30	35	4.6	915	21	AAC66561	Rat FGFR-1 flagged
	31	35	4.6	1983	22	AAD16027	Mouse soluble high
	32	35	4.6	2526	24	ABA01689	Mouse fibroblast g
	33	35	4.6	1503900	22	AAK95240	Human neuregulin-1
	34	35	4.6	1503900	22	AAK96733	Human neuregulin-1
	35	34	4.4	718	23	AAS77231	DNA encoding novel
	36	34	4.4	718	23	AAS77476	DNA encoding novel
	37	34	4.4	855	23	AAS55438	Streptococcus pneu
c	38	34	4.4	1628	21	AAC74342	Human secreted pro
c	39	33.4	4.4	193	21	AAC08794	Human secreted pro
	40	33.4	4.4	3734	23	AAS72406	DNA encoding novel
	41	33.4	4.4	3894	20	AAX55769	C. elegans Robo po
	42	33.4	4.4	3894	20	AAX57252	C. elegans ROBO cD
	43	33.4	4.4	5332	22	AAK52944	Human polynucleoti
	44	33.4	4.4	5489	22	AAK51960	Human polynucleoti
c	45	33.4	4.4	10684	17	AAT33758	Control region iso

ALIGNMENTS

RESULT 1
AAS28788/c
ID AAS28788 standard; cDNA; 1018 BP.
XX
AC AAS28788;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immunoglobulin encoding cDNA SEQ ID No 34.
XX
KW Immunoglobulin; primer; signal transduction pathway protein; cancer; ss;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
OS Homo sapiens.
XX
PN WO200155315-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01526.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215145.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231988.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234273.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457725/49.
DR P-PSDB; AAU18000.
XX
PT Isolated novel immunoglobulin polypeptide for monitoring the presence
PT and progression of diseases and for diagnosis -
XX
PS Claim 1, SEQ ID No 34; 551pp; English.
XX
CC Sequences AAS28765-AAS28877 represent cDNA molecules, which encode the
CC immunoglobulin polypeptides of the invention, and primers for the
CC polynucleotides. The polynucleotides and polypeptides can be used to
CC diagnose a pathological condition or a susceptibility to a pathological
CC condition in a subject by determining the presence or absence of a

```
CC   mutation in a DNA sequence or determining the presence or amount of
CC   expression of the protein. Alternatively the identification of a binding
CC   partner to a sequence allows determination of changes in protein
CC   activity. The sequences can be used as research tools for receptors or
CC   other signal transduction pathway proteins that interact with the
CC   polypeptides of the invention and can be used to treat, prevent or
CC   diagnose various types of disorders such as neurological disorders,
CC   cardiovascular disorders, gastrointestinal disorders, reproductive
CC   disorders, immune system disorders, renal disorders, muscular disorders,
CC   pulmonary disorders, proliferative disorders and cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1018 BP; 284 A; 225 C; 198 G; 311 T; 0 other;


  Query Match             73.0%;  Score 558.4;  DB 22;  Length 1018;
  Best Local Similarity   99.1%;  Pred. No. 9.8e-166;
  Matches  559;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      55 acaagttctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacct 114
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     963 AGAGGTTCTGTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCT 904

Qy     115 ggctcccaagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctc 174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     903 GGCTCCCAAGCATCTCTGATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTC 844

Qy     175 tggtaccgagaggaggggagagtggatttgaaatctgsaaacaamatcaattccagctct 234
           ||||||||||||||||||||||||||||||||||||:||||||:||||||||||||||||
Db     843 TGGTACCGAGAGGAGGGGAGAGTGGATTTGAAATCTGGAAACAAAATCAATTCCAGCTCT 784

Qy     235 gtctgtgtctcttccatcagtgaaaatgacaacggaatcagctttacctgcaggctgggg 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     783 GTCTGTGTCTCTTCCATCAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGG 724

Qy     295 agggatcagtccgtgtccgtttcggtggtgctgaatgttacttttcctcctctcctaagt 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     723 AGGGATCAGTCCGTGTCCGTTTCGGTGGTGCTGAATGTTACTTTTCCTCCTCTCCTAAGT 664

Qy     355 ggaaacgacttccaaacagttgaggaaggcagtaatgtgaagttggtttgcaatgtgaaa 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     663 GGAAACGACTTCCAAACAGTTGAGGAAGGCAGTAATGTGAAGTTGGTTTGCAATGTGAAA 604

Qy     415 gccaaccccccaggctcaaatgatgtggtacaaaaaacagtagtctcctcgatttagagaaa 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 GCCAACCCCCCAGGCTCAAATGATGTGGTACAAAAACAGTAGTCTCCTCGATTTAGAGAAA 544

Qy     475 agccgtcaccaaatccaacagacaagtgagtcttttcagctgtcaatcaccaaagtcgag 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 AGCCGTCACCAAATCCAACAGACAAGTGAGTCTTTTCAGCTGTCAATCACCAAAGTCGAG 484

Qy     535 aagcctgacaacggaacctacagttgtattgcaaagtcatctctgaaaacggagagcttg 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 AAGCCTGACAACGGAACCTACAGTTGTATTGCAAAGTCATCTCTGAAAACGGAGAGCTTG 424

Qy     595 gactttcacctgattgttaaagat 618
           ||||||||||||||||||||||| |
Db     423 GACTTTCACCTGATTGTTAAAGCT 400


RESULT   2
AAH35000
ID   AAH35000 standard; cDNA; 1958 BP.
XX
AC   AAH35000;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:2082.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   05-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG75595.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 3568-3569; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 1958 BP; 509 A; 480 C; 462 G; 495 T; 12 other;


  Query Match             57.6%;  Score 441;  DB 22;  Length 1958;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-128;
  Matches  441;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     325 ctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggc 384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     644 ctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggc 703

Qy     385 agtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtac 444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     704 agtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtac 763

Qy     445 aaaaacagtagtctcctcgatttagagaaaaagccgtcaccaaatccaacagacaagtgag 504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     764 aaaaacagtagtctcctcgatttagagaaaaagccgtcaccaaatccaacagacaagtgag 823

Qy     505 tcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtatt 564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     824 tcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtatt 883

Qy     565 gcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagataaaact 624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     884 gcaaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagataaaact 943

Qy     625 gtgggtgtaccaatagagcccattattgctgcatgtgttgtgatctttctgacattgtgc 684
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     944 gtqqqtqtaccaataqaqccattattgctqcatgtqttqtgatctttctqacattgtqc 1003

Qy     685 tttqqactqattgctaqaaqaaaqaaaataatqaaqctctqcatgaaggataaagaccct 744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1004 tttggactgattgctagaaggaaagaaaataatgaaqctctgcatgaaggataaagaccct 1063

Qy     745 cacaqtqaaacaqctctatqa 765
          |||||||||||||||||||||
Db    1064 cacaqtqaaacaqctctatqa 1084


RESULT   3
AAS28883/c
ID   AAS28883 standard; DNA; 19415 BP
XX
AC   AAS28883;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immunoglobulin encoding genomic DNA SEQ ID No 245
XX
KW   Immunoglobulin; signal transduction pathway protein; cancer; ds;
KW   antisense therapy, gene therapy, neurological disorder, renal disorder,
KW   cardiovascular disorder, gastrointestinal disorder, pulmonary disorder,
KW   reproductive disorder, immune system disorder, proliferative disorder,
KW   muscular disorder.
XX
OS   Homo sapiens.
XX
PN   WO200155315-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001426.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215145.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
```

```
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-457725/49.
XX
PT   Isolated novel immunoglobulin polypeptide for monitoring the presence
PT   and progression of diseases and for diagnosis -
XX
PS   Claim 1; SEQ ID No 245; 551pp; English.
XX
CC   Sequences AAS28878-AAS28926 represent genomic DNA molecules which encode
CC   the immunoglobulin polypeptides of the invention. The polynucleotides and
CC   polypeptides can be used to diagnose a pathological condition or a
CC   susceptibility to a pathological condition in a subject by determining
CC   the presence or absence of a mutation in a DNA sequence or determining
CC   the presence or amount of expression of the protein. Alternatively the
CC   identification of a binding partner to a sequence allows determination of
CC   changes in protein activity. The sequences can be used as research tools
CC   for receptors or other signal transduction pathway proteins that interact
CC   with the polypeptides of the invention and can be used to treat, prevent
CC   or diagnose various types of disorders such as neurological disorders,
CC   cardiovascular disorders, gastrointestinal disorders, reproductive
CC   disorders, immune system disorders, renal disorders, muscular disorders,
CC   pulmonary disorders, proliferative disorders and cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 19315 BP; 5674 A; 4041 C; 4034 G; 5565 T; 1 other;

  Query Match             37.1%;  Score 284;  DB 22;  Length 19315;
  Best Local Similarity   91.2%;  Pred. No. 2e-78;
  Matches  299;  Conservative    2;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      12 aagatttcttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaac 71
           | | | | | ||||   ||  |  ||   | |  | |    |  || ||||||||||||||
Db   14845 AGGTTGTTTACTCTGTATATCTAAATAGTTTCTCCCTATACTCCCAGGTTCTGTTTTAAC 14786

Qy      72 tgtgaatggtaaaactgagaactatatcctggatactacacctggctcccaagcatctct 131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14785 TGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCTGGCTCCCAAGCATCTCT 14726

Qy     132 gatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggg 191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   14725 GATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTCTGGTACCGAGAGGAGGG 14666

Qy     192 gagagtggatttgaaatctgsaaacaamatcaattccagctctgtctgtgtctttccat 251
           |||||||||||||||||||:|||||:||||||||||||||||||||||||||||||||
Db   14665 GAGAGTGCATTTGAAATCTGGAAACAAAATCAATTCCAGCTCTGTCTGTGTCTTTCCAT 14606

Qy     252 cagtgaaaaatgacaacggaatcagctttacctgcaggctgggggagggatcagtccgtgtc 311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14605 CAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGGAGGGATCAGTCCGTGTC 14546

Qy     312 cgtttcggtggtgctgaatgttacttttt 339
           |||||||||||||||||||||||||||||
Db   14545 CGTTTCGGTGGTGCTGAATGTTACTTGT 14518


RESULT   4
AAS37585
ID   AAS37585 standard; cDNA; 338 BP.
XX
AC   AAS37585;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #643.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A,  Lamson G;
PI   Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones WL,  Stache-Crain B;
XX
DR   WPI; 2001-530177/58.
XX
PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -
XX
PS   Claim 1; Page 757; 1193pp; English.
XX
CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39338 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.
XX
SQ   Sequence 338 BP; 93 A; 69 C; 91 G; 85 T; 0 other;

  Query Match             36.3%;  Score 277.4;  DB 22;  Length 338;
  Best Local Similarity   98.2%;  Pred. No. 2.9e-77;
  Matches  278;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      55 acaagttctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacct 114
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 agaggttctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacct 115
```

```
Qy      115 qqctvvvaaqcatctctgatatgtgctgttcaaaaccacaccagagaggaaggaactgctc 174
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      116 qqctccaaqcatctctgatatgtgctgttcaaaaccacaccagagaggaaggaactgctc 175

Qy      175 tggtaccgagagaggaggggagagtggatttgaaatctgsaaacaamatcaattccagctct 234
            ||||||||||||||||||||||||||||||||||||:||||||:|||||||||||||||||
Db      176 tggtaccgagagaggaggggagagtggatttgaaatctggaaacaaaatcaattccagctct 235

Qy      235 gtctgtgtctcttccatcagtgaaaatgacaacggaatcagctttacctgcaggctgggg 294
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      236 gtctgtgtctcttccatcagtgaaaatgacaacggaatcagctttacctgcaggctgggg 295

Qy      295 agggatcagtccgtgtccgtttcggtggtgctgaatgttactt 337
            ||||||||||||||||||||||| ||||||||||||||||||
Db      296 agggatcagtccgtgtccgtttcggaggtgctgaatgttactt 338
```

RESULT 5
AAS48121
ID AAS48121 standard; cDNA; 361 BP.
XX
AC AAS48121;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #1179.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G,
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer
XX
PS Claim 1; Page 892; 1194pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 361 BP; 114 A; 63 C; 90 G; 94 T; 0 other;

```
  Query Match           29.7%;  Score 227.2;  DB 22;  Length 361;
  Best Local Similarity  99.1%;  Pred. No. 2.1e-61;
  Matches  226;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 atgcaaatgqqqaaqatttcttctcttaqtaattttatttctgccacgtqaqatqacaagt 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      133 atgcaaatgqqqaaqatttcttctcttaqtaattttatttctgccacgtqaqatqacaagt 192

Qy       61 tctgttttaactgtgaatqgtaaaactqaqaactatatcctqqatactacacctqqctcc 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      193 tctgttttaactgtgaatggtaaaactgaqaactatatcctqqatactacacctqqctcc 252

Qy      121 caaqcatctctgatatgtgctgttcaaaaccacaccagagaqqaaqaactgctctggtac 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      253 caaqcatctctgatatgtgctgttcaaaaccacaccagagaqqaaqaactgctctggtac 312

Qy      181 cgagaqqaqqqqaqaqtqgatttgaaatctqsaaacaamatcaattcc 228
            |||||||||||||||||||||||||||||:|||||:|||||||||
Db      313 cgagaggaqqqqaqagtggatttgaaatctqgaaacaaaatcaattcc 360
```

RESULT 6
AAS38280
ID AAS38280 standard; cDNA; 328 BP.
XX
AC AAS38280;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #1338.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 932; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 328 BP; 102 A; 62 C; 81 G; 83 T; 0 other;

```
  Query Match           29.5%;  Score 225.4;  DB 22;  Length 328;
  Best Local Similarity  97.8%;  Pred. No. 7.4e-61;
  Matches  226;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 atgcaaatgqqqaaqatttcttctcttaqtaattttatttctgccacgtqaqatqacaagt 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       97 atgcaaatgqqqaaqatttcttctcttaqtaattttatttctgccacgtqaqatqacaagt 156
```

```
Qy      61 tctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacctggctcc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     157 tctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacctggctcc 216

Qy     121 caagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     217 caagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtac 276

Qy     181 cgagaggaggggagagtggatttgaaatctgsaaacaamatcaattccagc 231
           ||| |||||||||||||||||||||||||||:||||||:||| ||||| ||
Db     277 cgacaggaggggagagtggatttgaaatctggaaacaaaatccattcccgc 327


RESULT   7
AAF65373
ID   AAF65373 standard, cDNA, 359 BP.
XX
AC   AAF65373;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1129.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 706; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 359 BP; 113 A; 64 C; 92 G; 89 T; 1 other;

  Query Match             29.2%;  Score 223.6;  DB 22,  Length 359,
  Best Local Similarity   98.7%;  Pred. No. 2.9e-60;
  Matches  223;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 atgcaaatgggaagatttcttctcttagtaattttatttctgccacgtgagatgacaagt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 atgcaaatgggaagatttcttctcttagtaattttatttctgccacgtgagatgacaagt 192

Qy      61 tctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacctggctcc 120
           ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db     193 tctgttttaactgtgaatcgtaaaactgagaactatatcctggatactacacctggctcc 252

Qy     121 caagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 caagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtac 312

Qy     181 cgagaggaggggagagtggatttgaaatctgsaaacaamatcaatt 226
           ||||||||||||||||||||||||||||||:|||||:|||||||
Db     313 cgagaggaggggagagtggatttgaaatctggaaacaaaatcaatt 358


RESULT   8
AAF65369
ID   AAF65369 standard; cDNA; 359 BP.
XX
AC   AAF65369;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1125.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 705; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
```

```
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 359 BP; 106 A; 69 C; 95 G; 88 T; 1 other;


  Query Match             23.5%;  Score 179.8;  DB 22;  Length 359;
  Best Local Similarity   94.5%;  Pred. No. 1.9e-46;
  Matches  296;  Conservative    1;  Mismatches    9;  Indels    2;  Gaps    2;

Qy        1 atgcaaatgtggaaagatttcttctcttagtaatttttatttctgccacgtgagatgacaagt 60
            ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      133 atgcaaatgtggaaagatttcttctcttagtaattctatttctgccacgtgagatgacaagt 192

Qy       61 tctgttttaactgtgaatgtgtaaaactgagaactatatcctggatactaaacctggctcc 120
            |||| |||||||||||||| |||| ||||||||||||||||||||||||||||||||||||
Db      193 tctgttttaactgtgaatgtgtaaaactgagaactatatcctggatactaaacctggctcc 252

Qy      121 caagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtac 180
            |||||||||||||||||||||||||||| ||||||||||||||||| ||||||||||||
Db      253 caagcatctctgatatgtgctgttccaaaccacaccagagaggaag-actgctctggtac 311

Qy      181 cgagaggagggagagtgaatttgaaatctgaaacaa 218
            |||||||| ||||||| |||||||| |||||:||  ||
Db      312 cgagagga-gggagagggaatttgaaatctgaaacaaa 348


RESULT   9
AAD00024
ID   AAD00024 standard; DNA; 2949 BP.
XX
AC   AAD00024;
XX
DT   24-JUL-2000  (first entry)
XX
DE   Degenerate DNA encoding follistatin-related protein zfsta.
XX
KW   Follistatin-related protein; zfsta2; Transforming growth factor-beta;
KW   TGF-beta family; acute inflammatory condition; bursitis; myositis;
KW   chronic inflammatory demyelinating polyneuropathy; contact dermatitis;
KW   contact vulvovaginitis; sepsis; ulcerative colitis; dermatological;
KW   antiinflammatory; antiulcer; antibacterial; immunosuppressive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200022126-A1.
XX
PD   20-APR-2000.
XX
PF   05-OCT-1999;   99WO-US23179.
XX
PR   15-OCT-1998;   98US-0173044.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC,  Ellsworth JL;
XX
DR   WPI; 2000-317983/27.
XX
PT   New polypeptide comprising follistatin homology domain is useful for
PT   treating acute inflammatory conditions, bursitis, contact dermatitis,
PT   contact vulvovaginitis, myositis, sepsis and ulcerative colitis
XX
PS   Disclosure; Page 117-118; 125pp; English.
XX
CC   The present sequence encodes follistatin related protein zfsta2. This is
CC   a degenerate DNA that encompasses all DNAs that encode zfsta2 protein.
CC   zfsta2 gene and its protein are useful for binding to members of the
CC   Transforming growth factor (TGF)-beta family and mediating central
CC   nervous system, reproductive, haematopoietic and bone-related activities.
CC   This is useful for treating acute inflammatory conditions, bursitis,
CC   chronic inflammatory demyelinating polyneuropathy, various forms of
CC   contact dermatitis, contact vulvovaginitis, myositis, sepsis and
CC   ulcerative colitis. zfsta2 cDNA is also useful as a primer and probe for
CC   identifying genes encoding zfsta2 like proteins.
XX
SQ   Sequence 2949 BP; 614 A; 274 C; 477 G; 396 T; 1188 other;


  Query Match              5.4%;  Score 41.6;  DB 21;  Length 2949;
  Best Local Similarity   33.4%;  Pred. No. 0.025;
  Matches  103;  Conservative   46;  Mismatches  153;  Indels    6;  Gaps    2;

Qy      266 acggaatcagcttcacctgcaggctgggagaggagtcagtccgtgtccgtttcggtggtgc 325
            |:|     | :| :|| ||:      |   |  |: |:   || | :    |     |
Db     1352 aygtnggnaaytayacntgytaygcngaygggntaygarcargtntaycuraencayatht 1411

Qy      326 tgaatgttacttttcctcctcctcaagtggaaaac---gacttccaaacagttgaggaaag 382
            |  | || | : | || ||  |  |::| |   |:   || :  ||: |    ||:   |
Db     1412 tycargtnaaygtncencngtnathmgngtnta;cengarwsncargcnmgngarccng 1471

Qy      383 gcagtaatgtgaagttggtttgcaatgtgaaagccaacccccaggctcaaatgatgtggt 442
            |    | | :  :|    ||: |:|   |:|  | :||  |: | ||: |    |||:
Db     1472 gngtnacngcnwsnytnmgntgycaygcngarggnatheenaarecncarytnggntggy 1531

Qy      443 acaaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagaccagtg 502
              ||:||: | | :  : |:      |:   ::    || :  |  :  ||: | | :|
Db     1532 tnaaraaygggnathgayathacncenaarytnwsnaarcarytnacnytncargcnaayg 1591

Qy      503 agtctttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgta 562
              ::     |: |    ||::: || ||     |    ||:|  ||  | ||:|  ||:|
Db     1592 gnwsng   argtncayathwsnaaygtnmgntaygargayacngggngcntayacntgya 1648

Qy      563 ttgcaaag 570
            |:|| ||:
Db     1649 thgcnaar 1656


RESULT  10
AAV70230
ID   AAV70230 standard; DNA; 4092 BP.
XX
AC   AAV70230;
XX
DT   11-FEB-1999  (first entry)
XX
DE   Human receptor tyrosine kinase ROR 2 encoding DNA.
XX
KW   Receptor tyrosine kinase; Ror 1; Ror 2; Ptk 1; Ptk-2; detection;
KW   neurotrophin activity; trkB; proto-oncogene; tyrosine kinase receptor;
KW   binding protein; BDNF; NT-3; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             200..3031
FT                   /*tag=  a
XX
PN   US5843749-A.
XX
PD   01-DEC-1998.
XX
PF   06-JUN-1995;   95US-0469537.
XX
PR   17-MAR-1995;   95US-0406247.
PR   26-JUL-1991;   91US-0736559.
PR   28-OCT-1993;   93US-0144992.
PR   06-JUN-1995;   95US-0469537.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Maisonpierre PC,  Masiakowski P,  Yancopoulos GD;
XX
DR   WPI; 1999-044584/04.
DR   P-PSDB; AAW83168.
XX
```

```
PT   DNA encoding receptor tyrosine kinase proteins - and corresponding
PT   proteins
XX
PS   Example; Fig 27; 194pp; English.
XX
CC   The present invention describes nucleic acid molecules for ror-1,
CC   ror-2, ehk-1 and ehk-2. Also described are the corresponding proteins:
CC   Ror-1; Ror-2; Ehk-1; and Ehk-2. The proteins are orphan receptor
CC   tyrosine kinases. The present sequence encodes human ROR-2.
XX
SQ   Sequence 4092 BP; 884 A; 1229 C; 1171 G; 808 T; 0 other;


  Query Match             5.1%;  Score 39.2;  DB 20;  Length 4092;
  Best Local Similarity   46.9%;  Pred. No. 0.17;
  Matches  122;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy     325 ctgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggc 384
           ||||| | |   ||||    | | ||      |    ||      ||  |||   |||||
Db     368 ctgaaaggttacttttctgaattttctggagccagtaaacaatatcaccattgtccaaggc 427

Qy     385 agtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtac 444
           |  |  |   ||  |||||| |||  ||  |||||  |   || |  ||  ||||
Db     428 cagacggcaattctgcactgcaaggtggcaggaaacccacccccctaacgtgcggtggcta 487

Qy     445 aaaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgag 504
           || ||   |  |  | |     |||    | ||   |  ||||  |||||    |
Db     488 aagaatgatgccccggtggtgcaggagccgcggcggatcatcatccggaagacagaatat 547

Qy     505 tcttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtatt 564
             ||  |  |||  ||||    |  | || | | | ||||  ||   ||||   ||  |
Db     548 ggttcacgactgcgaatccaggacctggacacgacagacactggctactaccagtgcgtg 607

Qy     565 gcaaagtcatctctgaaaac 584
           || |         |||| ||
Db     608 gccaccaacgggatgaagac 627


RESULT  11
AAH98595
ID   AAH98595 standard; cDNA; 4710 BP.
XX
AC   AAH98595;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Human EST-derived coding sequence SEQ ID NO: 452.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02687.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   17-JUL-2000; 2000US-0617746.
PR   03-AUG-2000; 2000US-0631451.
PR   15-SEP-2000; 2000US-0663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Qian XB,  Wang Z,  Chen R,  Asundi V;
PI   Cao Y,  Drmanac RA,  Zhang J,  Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM23936.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Claim 1; Page 500-501; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention.
XX
SQ   Sequence 4710 BP; 1179 A; 1258 C; 1167 G; 1106 T; 0 other;


  Query Match             5.1%;  Score 39.2;  DB 22;  Length 4710;
  Best Local Similarity   51.1%;  Pred. No. 0.19;
  Matches   92;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy     391 gtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtacaaaaac 450
           || |   |||| ||| |||   ||| |  |||  || |    ||||  |||   ||  |
Db     836 gtcaccctggtgtgcgatgccgaaggcttcccagagcccaccatgagctggacaaaggat 895

Qy     451 agtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtctttt 510
             |    |   | ||   |||| ||  |  | |  |   |      |     ||   ||
Db     896 ggggaacagatagagcaagaggaagacgatgagaagtacatcttcagcgacgatagttcc 955

Qy     511 cagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaag 570
           ||||||  | ||||  || ||  || |||   ||| | |      ||||  || |||||  ||
Db     956 cagctgaccatcaaaaaggtggataagaacgacgaggctgagtacatctgcattgctgag 1015


RESULT  12
AAH98610
ID   AAH98610 standard; cDNA; 4710 BP.
XX
AC   AAH98610;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Human EST-derived coding sequence SEQ ID NO: 467.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25 JAN 2001; 2001WO-US02687.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   17-JUL-2000; 2000US-0617746.
PR   03-AUG-2000; 2000US-0631451.
PR   15-SEP-2000; 2000US-0663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Qian XB,  Wang Z,  Chen R,  Asundi V;
PI   Cao Y,  Drmanac RA,  Zhang J,  Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM23951.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
```

```
XX
PS   Claim 1; Page 509-510; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention.
XX
SQ   Sequence 4710 BP; 1179 A; 1258 C; 1167 G; 1106 T; 0 other;


  Query Match             5.1%;  Score 39.2;  DB 22;  Length 4710;
  Best Local Similarity   51.1%;  Pred. No. 0.19;
  Matches   92;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy     391 gtgaaqttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtacaaaaac 450
           || |   |||| ||| |||   ||| |  |||  || |    ||||  |||   ||  |
Db     836 gtcaccctggtgtgcgatgccgaaggcttcccagagcccaccatgagctggacaaaggat 895

Qy     451 agtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtctttt 510
             |    |   | ||   |||| ||  |  | |  |   |     |    ||   ||
Db     896 ggggaacagatagagcaagaggaagacgatgagaagtacatcttcagcgacgatagttcc 955

Qy     511 cagctgtcaatcaccaaaqtcqagaagcctqacaacqgaacctacagttgtattgcaaag 570
           |||||| | ||||  || || || |||   ||| | |     ||||  || |||||  ||
Db     956 cagctgaccatcaaaaagytggataagaacgacgaggctgagtacatctgcattgctgag 1015


RESULT  13
AAD14203
ID   AAD14203 standard; DNA; 8513 BP.
XX
AC   AAD14203;
XX
DT   06-NOV-2001  (first entry)
XX
DE   MESVR/EGFP/IRESNCAMPro(ori) vector.
XX
KW   Transcriptional regulatory element; translational regulatory element;
KW   gene expression; diagnostic application; therapeutic application;
KW   MESVR; retrovirus vector; enhanced green fluorescent protein; EGFP; IRES;
KW   internal ribosome entry sequence, N CAM, neural cell adhesion molecule,
KW   ori; origin of replication; SV40; human; mouse, ds.
XX
OS   Chimeric - Murine embryonic stem cell virus.
OS   Chimeric - Rous sarcoma virus.
OS   Chimeric - Mastadenovirus.
OS   Chimeric - Mus sp.
OS   Chimeric - Rhesus macaque polyoma virus.
OS   Chimeric - Homo sapiens.
OS   Chimeric - Synthetic.
XX
PN   WO200155371-A1.
XX
PD   02-AUG-2001.
XX
PF   26-JAN-2001; 2001WO-US02733.
XX
PR   28-JAN-2000; 2000US-0178816.
PR   02-MAR-2000; 2000US-0186496.
PR   30-MAY-2000; 2000US-0207804.
PR   07-SEP-2000; 2000US-0230852.
PR   07-SEP-2000; 2000US-0230956.
PR   12-JAN-2001; 2001US-0261312.
XX
PA   (SUKI ) SCRIPPS RES INST.
XX
PI   Mauro VP, Edelman GM, Chappell GM, Jones FS, Owens G, Meech R;
XX
DR   WPI; 2001-483242/52.
XX
PT   Identifying oligonucleotides with transcriptional/translational
PT   regulatory activity in eukaryotic cells by integrating an
PT   oligonucleotide into cell genome and detecting a change in expression
PT   of expressible polynucleotides -
XX
PS   Claim 35; Page 131-135; 172pp; English.
XX
CC   The present invention relates to a method for the identification of an
CC   oligonucleotide with transcriptional/translational regulatory activity in
CC   a eukaryotic cell. The method involves integrating an oligonucleotide
CC   into the genome of a eukaryotic cell so that it is linked to an
CC   expressible polynucleotide or contacting the eukaryotic cell with library
CC   of vectors, obtained by cloning a library of oligonucleotide into
CC   multiple copies of expression vectors comprising polynucleotide, and
CC   detecting changes in expression of the polynucleotide. The method is
CC   useful for identifying an oligonucleotide having transcriptional or
CC   translational activity in a eukaryotic cell. It is also useful for
CC   identifying synthetic transcriptional or translational regulatory
CC   elements. The transcriptional or translational regulatory elements are
CC   useful in a variety of gene expression configurations for regulating
CC   control of expression, and in expression vectors for controlling gene
CC   expressions in diagnostic and therapeutic applications. The present
CC   sequence is MESVR/EGFP/IRESNCAMPro(ori) vector based on
CC   Murine embryonic stem cell virus (MESV) retrovirus. The vector comprises
CC   nucleotide sequences encoding enhanced green fluorescent protein
CC   (EGFP) and human neural cell adhesion molecule (N-CAM) linked by an
CC   internal ribosome entry sequence (IRES). The upstream long terminal
CC   repeat (LTR) region of the retrovirus was modified to contain
CC   Rous sarcoma virus (RSV) enhancer elements. The downstream LTR U3 region
CC   was modified to contain a cassette containing a polylinker for the
CC   insertion of random oligonucleotides, adenovirus major late promoter,
CC   the initiator sequence (Inr) from mouse terminal deoxynucleotidyl
CC   transferase gene and a complete R region. The vector also contains a
CC   simian virus 40 (SV40) origin of replication (ori) and is useful
CC   for identifying synthetic transcriptional regulatory elements.
XX
SQ   Sequence 8513 BP; 2067 A; 2353 C; 2224 G; 1869 T; 0 other;


  Query Match             5.1%;  Score 39.2;  DB 22;  Length 8513;
  Best Local Similarity   51.1%;  Pred. No. 0.25;
  Matches   92;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy     391 gtgaaqttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtacaaaaac 450
           || |   |||| ||| |||   ||| |  |||  || |    ||||  |||   ||  |
Db    3607 gtcaccctggtgtgcgatgccgaaggcttcccagagcccaccatgagctggacaaaggat 3666

Qy     451 agtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaaagtgagtctttt 510
             |    |   | ||   |||| ||  |  | |  |   |     |    ||   ||
Db    3667 ggggaacagatagagcaagaggaagacgatgagaagtacatcttcagcgacgatagttcc 3726

Qy     511 cagctgtcaatcaccaaaqtcqagaagcctqacaacqgaacctacagttgtattgcaaag 570
           |||||| | ||||  || || || |||   ||| | |     ||||  || |||||  ||
Db    3727 cagctgaccatcaaaaagytggataagaacgacgaggctgagtacatctgcattgctgag 3786


RESULT  14
AAT90472
ID   AAT90472 standard; cDNA; 2869 BP.
XX
AC   AAT90472;
XX
DT   27-JAN-1998  (first entry)
XX
DE   Rat muscle-specific kinase (MuSK) cDNA.
XX
KW   Receptor tyrosine kinase; muscle specific kinase; MuSK; Dmk; rat;
KW   ligand; agrin; diagnosis; therapy; ss.
XX
OS   Rattus sp.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             135..2741
FT                   /*tag= a
FT   sig_peptide     135..191
FT                   /*tag= b
FT   mat_peptide     192..2738
FT                   /*tag= c
XX
PN   WO9721811-A2.
XX
PD   19-JUN-1997.
XX
PF   13-DEC-1996;   96WO-US20696.
XX
PR   10-MAY-1996;   96US-0644271.
PR   15-DEC-1995;   95US-0008657.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Bowen DC,  Glass DJ,  Valenzuela DM,  Yancopoulos GD;
XX
DR   WPI; 1997-332783/30.
DR   P-PSDB; AAW26610.
XX
PT   Nucleotide sequences encoding human agrin and muscle specific kinase
PT   and related receptor - used in diagnosis and treatment of disorder
PT   with muscle atrophy
XX
PS   Example 1; Fig 1; 120pp; English.
XX
CC   This cDNA sequence, deposited as ATCC 75498, includes a coding
CC   sequence for a novel rat receptor tyrosine kinase designated muscle
CC   specific kinase or MuSK (see AAW26610) that is expressed in normal
CC   and denervated muscle.  MuSK is alternatively referred to Dmk for
CC   denervated muscle kinase.  The cDNA clone was isolated from
CC   denervated muscle cDNA using primers (see AAT90476 and AAT90480)
CC   based on tyrosine kinase homology domains.  Human MuSK (see
CC   AAT90473) has also been identified.  Use of MuSK to generate anti-
CC   MuSK antibodies and in the diagnosis of neurological or other
CC   disorders is disclosed.  Assay systems that may be used to detect
CC   and/or measure ligands that bind the musk gene product are provided.
CC   A claimed method of promoting the growth, differentiation or
CC   survival of MuSK receptor-expressing cells involves administration
CC   to the cell of agrin (see AAW26609).  Such cells include muscle,
CC   heart, spleen, ovary and retina cells, or cells genetically
CC   engineered to express the MuSK receptor.
XX
SQ   Sequence 2869 BP; 727 A; 754 C; 726 G; 658 T; 4 other;
```

```
  Query Match             5.0%;  Score 38.6;  DB 18;  Length 2869;
  Best Local Similarity   55.3%;  Pred. No. 0.22;
  Matches   99;  Conservative    0;  Mismatches   74;  Indels    6;  Gaps    1;

Qy     391 gtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtacaaaaaac 450
           | ||  ||  | |||   ||| || || |||| ||| || ||||    ||| || |||
Db     267 gcgactttcatgtgcgccgtggaatcctaccctcagcctgaaatttcttggaccagaaat 326

Qy     451 agtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtctttt 510
           |  ! |||| || |  |     | | |||  ||   |||| | |||   || ||
Db     327 aaaattctcatcaagctgtttgacacccgctacagcatccgagagaacggtcag------ 380

Qy     511 cagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaa 569
           |  ||  | |||   |  || ||| |   |||  | || | |||| | || |  || ||
Db     381 ctcctcaccatcctgagtgtggaggacagtgatgatggcatctactgctgcacagccaa 439
```

```
RESULT  15
AAT87074
ID   AAT87074 standard; cDNA; 2869 BP.
XX
AC   AAT87074;
XX
```

```
DT   06-JAN-1998  (first entry)
XX
DE   Rat Dmk receptor cDNA.
XX
KW   Dmk receptor; tyrosine kinase receptor; signal transduction; assay;
KW   therapy; diagnosis; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             135..2741
FT                   /*tag= a
FT   sig_peptide     135..191
FT                   /*tag= b
FT   mat_peptide     192..2738
FT                   /*tag= c
XX
PN   US5656473-A.
XX
PD   12-AUG-1997.
XX
PF   21-JUL-1993;   93US-0095658.
XX
PR   19-JAN-1995;   95US-0374834.
PR   21-JUL-1993;   93US-0095658.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Rojas EA,  Valenzuela DM;
XX
DR   WPI; 1997-414593/38.
DR   P-PSDB; AAW26507.
XX
PT   New isolated human denervated muscle kinase receptor - used to
PT   develop products for the diagnosis and treatment of neurological,
PT   muscle or neuromuscular disorders
XX
PS   Example 1; Column 23-28; 31pp; English.
XX
CC   This cDNA clone, deposited as ATCC 75498,codes for rat Dmk receptor
CC   (AAW26507), a novel tyrosine kinase receptor that is expressed at high
CC   levels in denervated muscle.  It was isolated from a rat denervated
CC   skeletal muscle cDNA library following isolation of a partial
CC   sequence using primers (see AAT87075 and AAT87081) based on receptor
CC   tyrosine kinase (RTK) homology domains.  PCR primers (AAT87075-81)
CC   based on homology domains between novel Dmk and RTKs can be used to
CC   identify novel RTKs.  A human clone (see AAT87073) coding for a
CC   claimed human Dmk receptor (AAW26506) has been obtained.
XX
SQ   Sequence 2869 BP; 727 A; 754 C; 726 G; 658 T; 4 other;
```

```
  Query Match             5.0%;  Score 38.6;  DB 18;  Length 2869;
  Best Local Similarity   55.3%;  Pred. No. 0.22;
  Matches   99;  Conservative    0;  Mismatches   74;  Indels    6;  Gaps    1;

Qy     391 gtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtacaaaaaac 450
           | ||  ||  | |||   ||| || || |||| ||| || ||||    ||| || |||
Db     267 gcgactttcatgtgcgccgtggaatcctaccctcagcctgaaatttcttggaccagaaat 326

Qy     451 agtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagtctttt 510
           |  ! |||| || |  |     | | |||  ||   |||| | |||   || ||
Db     327 aaaattctcatcaagctgtttgacacccgctacagcatccgagagaacggtcag------ 380

Qy     511 cagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattgcaaa 569
           |  ||  | |||   |  || ||| |   |||  | || | |||| | || |  || ||
Db     381 ctcctcaccatcctgagtgtggaggacagtgatgatggcatctactgctgcacagccaa 439
```

Search completed: September 17, 2002, 10:01:52
Job time: 6020 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 10:04:48 ; Search time 56.43 Seconds
(without alignments)
432.513 Million cell updates/sec

Title: US-09-863-823-2
Perfect score: 1295
Sequence: 1 MQMGRFLLLVILFLPPEMTS RKKIMKLCMKDKDPHSETAL 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	176	13.6	853	1	IJBONC	neural cell adhesi
2	175	13.5	858	1	IJRTNC	neural cell adhesi
3	173.5	13.4	2783	2	T34416	hypothetical prote
4	171.5	13.2	946	1	A47299	ror-related recept
5	170	13.1	1896	2	T08851	Down syndrome cell
6	166	12.8	761	1	IJHUNG	neural cell adhesi
7	161	12.4	1011	2	T13669	neuromusculin - fr
8	160.5	12.4	1091	1	IJCHNL	neural cell adhesi
9	159	12.3	725	1	IJMSNG	neural cell adhesi
10	159	12.3	1115	1	IJMSNL	neural cell adhesi
11	156	12.0	1173	2	T25893	hypothetical prote
12	153.5	11.9	1344	2	T14316	rig-l protein - mo
13	153	11.8	2029	1	TDFFLK	protein tyrosine-p
14	152.5	11.8	5175	2	T20992	hypothetical prote
15	152.5	11.8	5198	2	T43290	hemicentin precurs
16	149	11.5	1447	2	A54100	tumor suppressor p
17	149	11.5	1912	2	A56178	protein-tyrosine-p
18	148.5	11.5	338	2	JC4776	limbic-system-asso
19	148.5	11.5	4391	2	A38096	perlecan precursor
20	146	11.3	7962	2	I38346	elastic titin - hu
21	145.5	11.2	773	2	T46283	hypothetical prote
22	144.5	11.2	338	2	JC5519	50K glycoprotein p
23	144.5	11.2	1348	2	S51656	vascular endotheli
24	144	11.1	458	1	WMMSR1	biliary glycoprote
25	144	11.1	521	2	JC1508	biliary glycoprote
26	142.5	11.0	458	2	JC1509	biliary glycoprote
27	142.5	11.0	521	2	S34338	biliary glycoprote
28	142	11.0	871	1	I48696	protein-tyrosine k
29	142	11.0	881	1	I48697	protein-tyrosine k
30	142	11.0	1894	2	C54689	protein-tyrosine-p
31	141	10.9	976	2	T29583	hypothetical prote
32	139	10.7	1450	2	A44027	165K myofibrillar
33	138.5	10.7	797	2	S38579	fibroblast growth
34	138	10.7	1880	2	T18531	tractin - medicina
35	137.5	10.6	1898	2	S46216	leukocyte antigen-
36	137	10.6	725	2	JE0100	neural cell adhesi
37	137	10.6	1092	1	JN0635	neural cell adhesi
38	136.5	10.5	458	2	S23969	cell-adhesion mole
39	136.5	10.5	458	2	S68177	C-CAM2a protein is
40	136.5	10.5	519	2	A44783	ecto-ATPase precur
41	136.5	10.5	1897	1	TDHULK	leukocyte antigen-
42	135	10.4	1273	2	T42405	sax-3 protein - Ca
43	135	10.4	1356	2	JC1402	protein-tyrosine k
44	134.5	10.4	1070	2	JC4593	protein-tyrosine k
45	134	10.3	1375	2	T13822	frazzled gene prot

ALIGNMENTS

RESULT 1
IJBONC
neural cell adhesion molecule short domain form precursor - bovine
N;Alternate names: NCAM-140
C;Species: Bos primigenius taurus (cattle)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 22-Jun-1999
C;Accession: A32976; A38778; B44290; S05402
R;Lipkin, V.M.; Khramtsov, N.V., Andreeva, S.G., Moshkyakov, M.V., Petukhova, G.V.; R
FEBS Lett. 254, 69-73, 1989
A;Title: Calmodulin-independent bovine brain adenylate cyclase. Amino acid sequence a
A;Reference number: A32976; MUID:89378239
A;Accession: A32976
A;Molecule type: mRNA
A;Residues: 1-853 <LIP>
A;Cross-references: GB:X16451; NID:g60; PIDN:CAA34470.1; PID:g61
A;Accession: A38778
A;Molecule type: protein
A;Residues: 20-35;51-61;113-117;122-147;155-161;262-275;279-302;353-360;369-382;544-5
A;Note: the authors identified this protein as calmodulin-independent adenylate cycla
R;Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
A;Title: Structural and immunological characterization of the amino-terminal domain o
A;Reference number: A44290; MUID:86140120
A;Accession: B44290
A;Molecule type: protein
A;Residues: 20-35;51-61;113-117;122-147;155-161;262-275;279-302;353-360;369-382;544-5
A;Note: 23-Glu was also found
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM
C;Comment: Various forms of NCAM are produced by alternative splicing.
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; i
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-853/Product: neural cell adhesion molecule, short domain form #status experiment
F;20-719/Domain: extracellular #status predicted <EXT>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-288/Domain: immunoglobulin homology <IMM3>
F;261-270/Region: NCAM binding #status predicted
F;321-396/Domain: immunoglobulin homology <IMM4>
F;428-490/Domain: immunoglobulin homology <IMM5>
F;527-604/Domain: fibronectin type III repeat homology <FN3A>
F;633-693/Domain: fibronectin type III repeat homology <FN3B>
F;720-737/Domain: transmembrane #status predicted <TMM>
F;738-853/Domain: intracellular #status predicted <INT>
F;41-95,139-189,235-286,328-394,435-488/Disulfide bonds: #status predicted
F;222,314,346,432,458,487/Binding site: carbohydrate (Asn) (covalent); #status predict

Query Match 13.6%; Score 176; DB 1; Length 853;
Best Local Similarity 28.4%; Pred. No. 7.7e-06;

```
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      41 QASLIC----------AVQNHTREEELLWYREEGRVDLKSGNKINS----SSVCVS 82
           | :| |           : :| : ||:  | | | :  |     : |    ||: :
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEKASWTRPEKQETLDGHMVVRSHARVSSLTLK 384

Qy      83 SISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 138
           ||   | |  : |     :|:|   | |: | | : | | |       ||: | : | |
Db     385 SIQYTDAG-EYVCTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 440

Qy     139 ANPQAQMMWYKNSSLLDLEK-SKHQTQQISESFQLSTIKVEKPDNGTYSCIAKSSLKTES 197
           | | | : |:::  ||     |  :| |  :  | :|   : | | |:| | : :  ||
Db     441 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 500

Qy     198 LDFHLIVKDKTVGVP----IEPIIAACVVIF 224
           |:| |:  | |   |    :|| :   | |
Db     501 LEFILVQAD-TPSSPSIDQVEPYSSTAQVQF 530


RESULT   2
IJRTNC
neural cell adhesion molecule short domain form precursor - rat
N;Alternate names: NCAM-140
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 22-Jun-1999
C;Accession: S00846; B37795; I58136
R;Small, S. J.; Shill, C. F.; Santoni, M. J.; Akeson, R.
J. Cell Biol. 105, 2345-2345, 1987
A;Title: Identification of a cDNA clone that contains the complete coding sequence for a
A;Reference number: S00846; MUID:88059265
A;Accession: S00846
A;Molecule type: mRNA
A;Residues: 1-858 <SMA>
A;Cross-references: EMBL:X06564
R;Small, S.J.; Akeson, R.
J. Cell Biol. 111, 2089-2096, 1990
A;Title: Expression of the unique NCAM VASE exon is independently regulated in distinct
A;Reference number: A37795; MUID:91035620
A;Accession: B37795
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 440-481 <SM2>
R;Small, S.J.; Haines, S.L.; Akeson, R.A.
Neuron 1, 1007-1017, 1988
A;Title: Polypeptide variation in an N-CAM extracellular immunoglobulin-like fold is dev
A;Reference number: I58136; MUID:90166485
A;Accession: I58136
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 355-364 <RES>
A;Cross-references: GB:M32011; NID:g205643; PIDN:AAA11679.1; PID:g205644
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C;Comment: Various forms of NCAM are produced by alternative splicing.
C;Genetics:
A;Gene: NCAM
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; si
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-858/Product: neural cell adhesion molecule, short domain form #status predicted <MA
F;20-721/Domain: extracellular #status predicted <EXT>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-290/Domain: immunoglobulin homology <IMM3>
F;268-272/Region: NCAM binding #status predicted
F;323-398/Domain: immunoglobulin homology <IMM4>
F;430-492/Domain: immunoglobulin homology <IMM5>
F;529-606/Domain: fibronectin type III repeat homology <FN3A>
F;635-695/Domain: fibronectin type III repeat homology <FN3B>
F;722-739/Domain: transmembrane #status predicted <IMM>
F;740-858/Domain: intracellular #status predicted <INT>
F;41-96,139-189,235-288,330-396,437-490/Disulfide bonds: #status predicted
F;222,316,348,434,466,489/Binding site: carbohydrate (Asn) (covalent) #status predict

  Query Match             13.5%;  Score 175;  DB 1;  Length 858;
  Best Local Similarity   28.4%;  Pred. No. 9.3e-06;
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      41 QASLIC----------AVQNHTREEELLWYREEGRVDLKSGNKINS----SSVCVS 82
           | :| |           : :| : ||:  | | | :  |     : |    ||: :
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEKASWTRPEKQETLDGHMVVRSHARVSSLTLK 384

Qy      83 SISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 138
           ||   | |  : |     :|:|   | |: | | : | | |       ||: | : | |
Db     385 SIQYTDAG-EYVCTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 440

Qy     139 ANPQAQMMWYKNSSLLDLEK-SKHQTQQISESFQLSTIKVEKPDNGTYSCIAKSSLKTES 197
           | | | : |:::  ||     |  :| |  :  | :|   : | | |:| | : :  ||
Db     441 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 500

Qy     198 LDFHLIVKDKTVGVP----IEPIIAACVVIF 224
           |:| |:  | |   |    :|| :   | |
Db     501 LEFILVQAD-TPSSPSIDQVEPYSSTAQVQF 530


RESULT   3
T34416
hypothetical protein F12F3.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34416
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F12F3.
A;Reference number: Z21521
A;Accession: T34416
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2783 <FUL>
A;Cross-references: EMBL:U80022; PIDN:AAC25886.1; GSPDB:GN00023; CESP:F12F3.2
A;Experimental source: strain Bristol N2, clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.2
A;Map position: 5
A;Introns: 45/3; 98/1; 451/3; 509/1; 2314/3; 2341/3; 2378/3; 2414/1; 2453/3; 2474/2;

  Query Match             13.4%;  Score 173.5;  DB 2;  Length 2783;
  Best Local Similarity   24.1%;  Pred. No. 5e-05;
  Matches   60;  Conservative   36;  Mismatches  100;  Indels   53;  Gaps    9;

Qy       1 GNELLLVILELPKEMISSVLLVNGKIENYILDIIP----------------------GS 40
           | |  |:   | :   |  :|:  |  :  | :||                      |
Db    1562 GTFNCLIENELGQASASCQVTIFNKPAS--IQSTPDHSLERNLVPTLQKALNNESAQAGQ 1619

Qy      41 QASLICAVQNHTREEELLWYREE------GRVDLKSGNKINSSSVCVSSISENDNGISFT 94
           | | | : : :  : |::::      || :| |  | |   ||  ::   | |  :
Db    1620 QIMLDCEISSRS-PSIVAWERDDRRIESAGRYELSSDKKSNHKLVC-HAVQSQDIG-KYR 1676

Qy      95 C----RLGRDQS---VSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMW 147
           |    : |  :|   |:|  |     |    :     |   | |: | | | | :|  : |
Db    1677 CVVINAYGYAFSECNVAVEIVIKEIAPSESATISDSTATIGHNITLECMVFGSPAPEVSW 1736

Qy     148 YKNSSLLDLEKSKHQTQQISESFQLSTIKVEKPDNGTYSCIAKS----------SLKT 195
           |:   :   :  | |  : : :|||:| |  | | | | | |             ||
Db    1737 IKNGEKISIFKRIRQTQDENGNCKLSISKAESDDMGVYVCSAISVAGVDSISSMVMIAKT 1796

Qy     196 ESLDFHLIV 204
             | |||::
Db    1797 TGTDSHLVI 1805


RESULT   4
A47299
```

```
ror-related receptor RTK - Pacific electric ray
N;Contains: protein-tyrosine kinase (EC 2.7.1.112)
C;Species: Torpedo californica (Pacific electric ray)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-May-2000
C;Accession: A47299
R;Jennings, C.G.; Dyer, S.M.; Burden, S.J.
Proc. Natl. Acad. Sci. U.S.A. 90, 2895-2899, 1993
A;Title: Muscle-specific trk-related receptor with a kringle domain defines a distinct c
A;Reference number: A47299; MUID:93219391
A;Accession: A47299
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-946 <JEN>
A;Cross-references: GB:L11311; NID:g290857; PIDN:AAA49285.1; PID:g290858
A;Experimental source: electric organ
A;Note: sequence extracted from NCBI backbone (NCBIN:128724, NCBIP:128726)
C;Superfamily: Torpedo ror-related receptor; immunoglobulin homology; kringle homology;
C;Keywords: ATP, glycoprotein, kringle, phosphotransferase, transmembrane protein, tyros
F;44-103/Domain: immunoglobulin homology <IM1>
F;137-195/Domain: immunoglobulin homology <IM2>
F;229-287/Domain: immunoglobulin homology <IM3>
F;464-542/Domain: kringle homology <KR3>
F;572-588/Domain: transmembrane #status predicted <TMN>
F;650-940/Domain: protein kinase homology <KIN>
F;658-666/Region: protein kinase ATP-binding motif
F;225,340,477,544/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             13.2%;  Score 171.5;  DB 1;  Length 946;
  Best Local Similarity   25.1%;  Pred. No. 2e-05;
  Matches   63;  Conservative   38;  Mismatches   95;  Indels   55;  Gaps   10;

Qy       7 LLLVIL---------FLPR--EMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREE 55
           ||:: |          ||:  ::|| : ||:   |          :|| :||| ::
Db      10 LLMIFLVTTGGSADGILPKAPQITSPLETVDALVE---------EEASFMCAVDSYP-AA 59

Qy      56 ELLWYREEGRV---DLKSGNKINSSSVCVSSISENDNGI-SFTCRLGRDQSVSVSVVLNV 111
           |: | |    :   | :   | |   : : |: : |||:   |   |   |      | |
Db      60 EITWTRNNIPIRPFDTRYSTKENGQILTILSVEDTDNGVYCCTANNGMGSSAQSCGALQV 119

Qy     112 TFPP--LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSES 169
             |  :    | : :  || | | |:   ||:  : |:|: : |  :: |  :    ||
Db     120 KMKPKIIRPPTDVRAL-LGSKVVLPCSTMGNPKPAISWFKDETALKNDQPRTSV---LES 175

Qy     170 FQLSITKVEKPDNGTYSCIAKSSLKTE-----------------------SLDFHLIVK 205
             | |  |:  | | | |:|::||  |                       |    :|::
Db     176 GNLRIRNVQLEDAGKYRCLARNSLGFEYSRSAALEVQVSARIVKAPTSQNVSYCSEVILQ 235

Qy     206 DKTVGVPIEPI 216
           |  | ||  |
Db     236 CKATGFPIPTI 246


RESULT   5
T08851
Down syndrome cell adhesion protein 1 - human (fragment)
N;Alternate names: Down syndrome cell adhesion molecule
C;Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08851
R;Yamakawa, K.; Huo, Y.K.; Haendel, M.A., Hubert, R., Chen, X.N., Lyons, G.E., Korenberg
submitted to the EMBL Data Library, September 1997
A;Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down sy
A;Reference number: Z16495
A;Accession: T08851
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1896 <YAM>
A;Cross-references: EMBL:AF023449; NID:g3169765; PID:g3169766
A;Experimental source: brain; developmental stage: 14 weeks; fetal
C;Genetics:
A;Gene: DSCAM
A;Map position: 21q22
A;Note: derived from alternately-spliced mRNA
C;Function:
A;Description: involved in nervous system development
C;Keywords: alternative splicing


  Query Match             13.1%;  Score 170;  DB 2;  Length 1896;
  Best Local Similarity   29.8%;  Pred. No. 5.9e-05;
  Matches   57;  Conservative   31;  Mismatches   69;  Indels   34;  Gaps   10;

Qy      15 PREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKSGNK- 73
           ||:: |||                 ||| || |:|   | ::|| |||  | : |  |
Db     306 PRKVKSSV---------------GSQVSLSCSVTG-TEDQELSWYR-NGEI-LNPGKNV 346

Qy      74 ----INSSSVCVSSISENDNGISFTCRLGRDQSVS---VSVVLNVTFPPLLSGNDFQTVE 126
               ||  :: :  : ::| | :: | : :|:  :   | |||    | ::|    : |
Db     347 RITGINHENLIMDHMVKSDGG-AYQCFVRKDKLSAQDYVQVVLEDGTPKIISAFSEKVVS 405

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQ--TSES---FQLSITKVEKPD 181
             | |:|||  |   : |  :   : |:   |:| |  |||      |:|:  :  |
Db     406 PAEPVSLMCNVKGTPLPTITWTLDDDFI LKGGSHRISQMITSEGNVVSYLNISSSQVRD 464

Qy     182 NGTYSCIAKSS 192
           | | | | :|
Db     465 GGVYRCTANNS 475


RESULT   6
IJHUNG
neural cell adhesion molecule 1 GPI-anchored splice form precursor, muscle-specific -
N;Alternate names: CD56; NCAM-120
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 28-Jan-2000
C;Accession: S07784; A26883
R;Barton, C.H.; Dickson, G.; Gower, H.J.; Rowett, L.H.; Putt, W.; Elsom, V.; Moore, S
Development 104, 165-173, 1988
A;Title: Complete sequence and in vitro expression of a tissue-specific phosphatidyli
A;Reference number: S07784; MUID:89305258
A;Accession: S07784
A;Molecule type: mRNA
A;Residues: 1-761 <BAR>
A;Cross-references: EMBL:X16841; NID:g35005; PIDN:CAA34739.1; PID:g35006
R;Dickson, G.; Gower, H.J.; Barton, C.H.; Prentice, H.M.; Elsom, V.L.; Moore, S.E.; C
Cell 50, 1119-1130, 1987
A;Title: Human muscle neural cell adhesion molecule (N-CAM): identification of a musc
A;Reference number: A90895; MUID:87301755
A;Accession: A26883
A;Molecule type: mRNA
A;Residues: 491-761 <DIC>
A;Cross-references: GB:M17409; NID:g189097; PIDN:AAA59912.1; PID:g386979
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM
C;Comment: Various forms of NCAM are produced by alternative splicing.
C;Genetics:
A;Gene: GDB:NCAM1; NCAM; CD56
A;Cross-references: GDB:119448; OMIM:116930
A;Map position: 11q22.2 11q22.3
C;Superfamily: neural cell adhesion molecule, fibronectin type III repeat homology; i
C;Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membra
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-761/Product: neural cell adhesion molecule phosphatidylinositol-linked form, mus
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-285/Domain: immunoglobulin homology <IMM3>
F;263-272/Region: NCAM binding #status predicted
F;322-387/Domain: immunoglobulin homology <IMM4>
F;419-481/Domain: immunoglobulin homology <IMM5>
F;499-587/Domain: fibronectin type III repeat homology #status atypical <FN3A>
F;633-720/Domain: fibronectin type III repeat homology #status atypical <FN3B>
F;41-96,139-189,235-287,329-385,426-479/Disulfide bonds: #status predicted
F;222,315,347,423,449,478/Binding site: carbohydrate (Asn) (covalent) #status predict
```

```
  Query Match             12.8%;  Score 166;  DB 1;  Length 761;
  Best Local Similarity   28.0%;  Pred. No. 4e-05;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      41 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 86
           | :| |                : :|  : ||: |    :| : ::|  ::  ||: : ||
Db     424 QVTLTCEASGDPIPSITWRTSTRNISSEEKTL---DGHMVVRSHARV--SSLTLKSIQY 477

Qy      87 NDNGISFIC----RIGDRQSVSVSVVLNVIFPPLLSGNDFQIVEEGSNVKLVCNVKANPQ 142
           | |  : |     :|:|   | |: | | : | | |       ||: | : | | | |
Db     478 TDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYIWEGNQVNITCEVFAYPS 533

Qy     143 AQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSTIRVEKPDNGIYSCIAKSSLNIESLDFH 201
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     534 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 593

Qy     202 LIVKDKTVGVP-  -IEPIIAACVVIF 224
           |:  | |   |    :|| :   | |
Db     594 LVQAD-TPSSPSIDQVEPYSSTAQVQF 619


RESULT   7
TI4669
neuromusculin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13 Aug 1999 #text_change 17 Nov 2000
C;Accession: TI3669
R;Kania, A.; Han, P.L.; Kim, Y.T.; Bellen, H.
Neuron 11, 673-687, 1993
A;Title: Neuromusculin, a Drosophila gene expressed in peripheral neuronal precursors an
A;Reference number: Z17697; MUID:94000831
A;Accession: TI3669
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1011 <KAN>
A;Cross-references: EMBL:L28146; NID:g385073; PID:g385074; PIDN:AAA03750.1
C;Genetics.
A;Gene: nrm
A;Cross-references: FlyBase:FBgn0005629


  Query Match             12.4%;  Score 161;  DB 2;  Length 1011;
  Best Local Similarity   24.3%;  Pred. No. 0.00014;
  Matches   46;  Conservative   34;  Mismatches   81;  Indels   28;  Gaps    6;

Qy      23 LTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-----EEGRVDL--------- 68
           : :||   :    |  ||:  | | :       | ||        |  ::
Db     265 IDINGVKHH     FVQGSKVVLPCDFIGCAEPAVNIIWYNETLLINSSGENLTFVRSFSL 320

Qy      69 KSGNKINSSSVCVSSISENDNGISFIC--------RIGDRQSVSVSVVLNVIFPPLLSGND 121
           ||    :: |  : : :  :|   | |         :: ||: :| :: | | :|||::  :
Db     321 KSDGTFHTQSFLIENATREFNDPVFPCFAFNIVIQINREKPISSALTLEVLYPPVVKVSP 380

Qy     122 FQIVEEGSNVKLV-CNVFANPPA--QMMWYKNSSLLDLEKSRHQIQQTSESFQLSTIRVE 178
               | : |: |   ||| :  |: ||:|  |:::  : |     ||:  | |   |
Db     381 SAITANISEIVLLNCEYFANPASLTQVEWYRNDILVNVNDTTHYKGGNSENVALVIKSTE 440

Qy     179 KPDNGIYSC 187
           | | | |||
Db     441 KEDIGNYSC 449


RESULT   8
I.ICHNL
neural cell adhesion molecule long domain form precursor - chicken
N;Alternate names: NCAM 180
N;Contains: neural cell adhesion molecule, short domain form (NCAM-140)
C;Species: Gallus gallus (chicken)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 22-Jun-1999
C;Accession: A43613; B43613; A25435; B25435; A46550; S36950; A44369; A60852; S29668
R;Cunningham, B.A.; Hemperly, J.J.; Murray, B.A.; Prediger, E.A.; Brackenbury, R.; Edelm
Science 236, 799-806, 1987
A;Title: Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell
A;Reference number: A43613; MUID:87206190
A;Accession: A43613
A;Molecule type: mRNA
A;Residues: 1-175 <CU2>
A;Cross-references: GB:M15860
A;Accession: B43613
A;Molecule type: protein
A;Residues: 20-44,120-127,202-221,320-342,399-415,640-659,822-828 <CUN>
A;Note: Asn-222 probably binds carbohydrate, Asn-226 probably does not
R;Hemperly, J.J.; Murray, B.A.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 3037-3041, 1986
A;Title: Sequence of a cDNA clone encoding the polysialic acid-rich and cytoplasmic d
A;Reference number: A25435; MUID:86206089
A;Accession: A25435
A;Molecule type: mRNA
A;Residues: 128-1091 <HEM>
A;Cross-references: GB:M13210
A;Accession: B25435
A;Molecule type: protein
A;Residues: 128-140;222-240;428-439;611-631;744-760;763-781;1089-1094 <HE2>
R;Murray, B.A.; Owens, G.C.; Prediger, E.A.; Crossin, K.L.; Cunningham, B.A.; Edelman
J. Cell Biol. 103, 1431-1439, 1986
A;Title: Cell surface modulation of the neural cell adhesion molecule resulting from
A;Reference number: A46550; MUID:87033934
A;Accession: A46550
A;Molecule type: DNA
A;Residues: 810-1070 <MUR>
A;Cross-references: GB:X04479
R;Sasner, M.; Covault, J.
submitted to the EMBL Data Library, February 1993
A;Reference number: S36950
A;Accession: S36950
A;Molecule type: DNA
A;Residues: 1-17 <SAS>
A;Cross-references: EMBL:X70342; NID:g417631; PIDN:CAA49807.1; PID:g417632
R;Colwell, G.; Li, B.; Forrest, D.; Brackenbury, R.
Genomics 14, 875-882, 1992
A;Title: Conserved regulatory elements in the promoter region of the N-CAM gene.
A;Reference number: A44369; MUID:93122797
A;Accession: A44369
A;Molecule type: DNA
A;Residues: 1-17 <COP>
A;Cross-references: EMBL:Z12128; NID:g63653; PIDN:CAA78113.1; PID:g63654
A;Experimental source: White Leghorn
R;Cole, G.J.; Loewy, A.; Cross, N.V.; Akeson, R.; Glaser, L.
J. Cell Biol. 103, 1739-1744, 1986
A;Title: Topographic localization of the heparin-binding domain of the neural cell ad
A;Reference number: A60852; MUID:87057627
A;Accession: A60852
A;Molecule type: protein
A;Residues: 20-29 <COL>
R;Rao, Y.; Wu, X.F.; Gariepy, J.; Rutishauser, U.; Siu, C.H.
J. Cell Biol. 118, 937-949, 1992
A;Title: Identification of a peptide sequence involved in homophilic binding in the n
A;Reference number: A43280; MUID:92364934
A;Contents: annotation; homophilic binding region
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM
C;Comment: Various forms of NCAM are produced by alternative splicing.
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; i
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1091/Product: neural cell adhesion molecule, long domain form #status experiment
F;20-809,1071-1091/Product: neural cell adhesion molecule, short domain form #status
F;20-711/Domain: extracellular #status predicted <EXT>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-289/Domain: immunoglobulin homology <IMM3>
F;262-271/Region: NCAM binding #status experimental
F;322-387/Domain: immunoglobulin homology <IMM4>
F;419-481/Domain: immunoglobulin homology <IMM5>
```

F;518-595/Domain: fibronectin type III repeat homology <FN3A>
F;624-685/Domain: fibronectin type III repeat homology <FN3B>
F;712-729/Domain: transmembrane #status predicted <TMM>
F;730-1091/Domain: intracellular #status predicted <INT>
F;41-96,139-189,235-287,329-385,426-479/Disulfide bonds: #status predicted
F;222/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;315,347,423,449,478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 12.4%; Score 160.5; DB 1; Length 1091;
Best Local Similarity 26.9%; Pred. No. 0.00017;
Matches 60; Conservative 37; Mismatches 87; Indels 39; Gaps 10;

```
Qy      25 VNGKTENYILDTTPGSQASLIC--------------AVQNHTREEELLWYREEGRVDLKS 70
           |  ||   : |     | :| |              : :| : ||: |    :||: ::|
Db     313 VENKTAMELED-----QITLTCEASGDPIPSITWKTSTRNISNEEKTL----DGRIVVRS 363

Qy      71 GNKINSSSVCVSSISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVE 126
             ::  ||: :  |   | |  : |     :|:|   | :: | | : | | |
Db     364 HARV--SSLTLKEIQYIDAG-EYVCTASNTIGQD---SQAMYLEVQYAPKLQGPVAVYTW 417

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTY 185
           ||: | : | | | | | : |::: {{     |  :|  |   :  | :|   : | | |
Db     418 EGNQVNITCEVFAYPSAVISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNY 477

Qy     186 SCIAKSSLKTESLDFHLIVKDKTVGVP----IEPIIAACVVIF 224
           :| | : :  || :| |:  | |   |    :||  :   | |
Db     478 NCTAVNRIGQESSEFILVQAD-TPSSPSIDRVEPYSSTARVEF 519
```

RESULT 9
IJMSNG
neural cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse
N;Alternate names: NCAM-120
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 31-Dec-2000
C;Accession: A29673; S00382; A44290
R;Barthels, D.; Santoni, M.J.; Wille, W ; Ruppert, C ; Chaix, J C ; Hirsch, M R ; Fonte
EMBO J. 6, 907-914, 1987
A;Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000
A;Reference number: A29673; MUID:87246524
A;Accession: A29673
A;Molecule type: mRNA
A;Residues: 1-725 <BAR>
A;Cross-references: EMBL:Y00051; NID:g53342; PIDN:CAA68263.1; PID:g53343
R;Barbas, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.
EMBO J. 7, 625-632, 1988
A;Title: Differential splicing and alternative polyadenylation generates distinct NCAM t
A;Reference number: S00382; MUID:88283628
A;Accession: S00382
A;Molecule type: DNA
A;Residues: 642-656,'D',658-725 <BA2>
A;Cross-references: EMBL:X07195
R;Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
A;Title: Structural and immunological characterization of the amino-terminal domain of m
A;Reference number: A44290; MUID:86140120
A;Accession: A44290
A;Molecule type: protein
A;Residues: 20-36 <ROU>
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C;Comment: Several forms of NCAM are produced by alternative splicing. See also PIR.IJMS
C;Genetics:
A;Gene: NCAM
A;Map position: 9
A;Introns: 701/1
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane
F;1-19/Domain: signal sequence #status predicted <SIG>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-290/Domain: immunoglobulin homology <IMM3>
F;263-272/Region: NCAM binding #status predicted
F;323-388/Domain: immunoglobulin homology <IMM4>
F;420-482/Domain: immunoglobulin homology <IMM5>
F;519-596/Domain: fibronectin type III repeat homology <FN3A>
F;625-685/Domain: fibronectin type III repeat homology <FN3B>
F;41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted
F;222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 12.3%; Score 159; DB 1; Length 725;
Best Local Similarity 27.5%; Pred. No. 0.00013;
Matches 57; Conservative 35; Mismatches 81; Indels 34; Gaps 9;

```
Qy      41 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 86
           | :| |               : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEQDL----DGHMVVRSHARV--SSLTLKSIQY 378

Qy      87 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 142
           | |  : |     :|:|   | ||: |   : | | |       ||: | : | | | |
Db     379 RDAG-EYMCTASNTIGQD---SQSIDLEFQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     143 AQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 201
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     202 LIVKDKTVGVP----IEPIIAACVVIF 224
           |:  | |   |    :||  :   | |
Db     495 LVQAD-TPSSPSIDRVEPYSSTAQVQF 520
```

RESULT 10
IJMSNL
neural cell adhesion molecule 1 precursor, long domain splice form - mouse
N;Alternate names: NCAM-180
N;Contains: neural cell adhesion molecule, short domain splice form (NCAM-140)
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 31-Dec-2000
C;Accession: A29673; S00844; S00384; A28281; A44290; S00383
R;Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fon
EMBO J. 6, 907-914, 1987
A;Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,
A;Reference number: A29673; MUID:87246524
A;Accession: A29673
A;Molecule type: mRNA
A;Residues: 1-548,'T',550-571,'T',573-574,'D',576-588,'MQPS',593,'S',595-599,'P',601,
A;Cross-references: EMBL:Y00051; NID:g53342; PIDN:CAA68263.1; PID:g53343
R;Santoni, M.J.; Barthels, D.; Barbas, J.A.; Hirsch, M.R.; Steinmetz, M.; Goridis, C.
Nucleic Acids Res. 15, 8621-8641, 1987
A;Title: Analysis of cDNA clones that code for the transmembrane forms of the mouse n
A;Reference number: S00844; MUID:88067687
A;Accession: S00844
A;Molecule type: mRNA
A;Residues: 529-809,1077-1115 <SAN>
A;Cross-references: EMBL:X06328; NID:g53322; PIDN:CAA29641.1; PID:g817984
R;Barbas, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.
EMBO J. 7, 625-632, 1988
A;Title: Differential splicing and alternative polyadenylation generates distinct NCA
A;Reference number: S00382; MUID:88283628
A;Accession: S00384
A;Molecule type: DNA
A;Residues: 642-1115 <BAR>
A;Cross-references: EMBL:X07195
R;Barthels, D.; Vopper, G.; Wille, W.
Nucleic Acids Res. 16, 4217-4225, 1988
A;Title: NCAM-180, the large isoform of the neural cell adhesion molecule of the mous
A;Reference number: A28281; MUID:88247737
A;Accession: A28281
A;Molecule type: mRNA
A;Residues: 804-1081 <BA3>
A;Cross-references: EMBL:X07244; NID:g53321; PIDN:CAA30230.1; PID:g929720
R;Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986

```
A;Title: Structural and immunological characterization of the amino terminal domain of N
A;Reference number: A44290; MUID:86140120
A;Accession: A44290
A;Molecule type: protein
A;Residues: 20-96 <BOG>
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C;Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:IJMS
C;Genetics:
A;Gene: NCAM
A;Map position: 9
A;Introns: 643/1; 701/1; 770/2; 809/2; 1076/2
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; si
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1115/Product: neural cell adhesion molecule, long domain splice form #status experi
F;20-809,1077-1115/Product: neural cell adhesion molecule, short domain splice form #sta
F;20-711/Domain: extracellular #status predicted <EXT>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-290/Domain: immunoglobulin homology <IMM3>
F;262-272/Region: NCAM binding #status predicted
F;323-388/Domain: immunoglobulin homology <IMM4>
F;420-482/Domain: immunoglobulin homology <IMM5>
F;519-596/Domain: fibronectin type III repeat homology <FN1A>
F;625-685/Domain: fibronectin type III repeat homology <FN3B>
F;712-729/Domain: transmembrane #status predicted <TMM>
F;730-[illegible]/Domain: extracellular #status predicted <INT>
F;41-96,139-189,235-288,330-386,427-480/disulfide bonds: #status predicted
F;222,316,348,424,452,479/binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match              12.3%;  Score 159;  DB 1;  Length 1115;
  Best Local Similarity    27.5%;  Pred. No. 0.00022;
  Matches   57;  Conservative   85;  Mismatches   81;  Indels   34;  Gaps    9;

Qy      41 QASLIC          --    AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 86
           | :| |               : :|  : ||: |    :|  : ::|  ::  ||: : ||
Db     425 QVTLTCEASGDPIPSITWRTSTRNISSEEQDL----DGHMVVRSHARV--SSITLKSIQY 478

Qy      87 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDEQIVEEGSNVKLVCNVKANPQ 142
           | |  : |      :|:|   | |: |   : | | |        ||: | : | | | |
Db     479 RDAG-EYMCTASNTIGQD---SQSIDLEFQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 534

Qy     143 AQMMWYKNSSLIDLEK-SRHQIQQTSFSFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 201
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     535 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 594

Qy     202 LIVKDKTVGVP----IEPIIAACVVIF 224
           |:  | |  |     :|| :   | |
Db     595 LVQAD-TPSSPSIDRVEPYSSTAQVQF 620


RESULT  11
T25893
hypothetical protein T19B4.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25893
R;Gattung, S.
submitted to the EMBL Data Library, November 1996
A;Description: The sequence of C. elegans cosmid T19B4.
A;Reference number: Z20106
A;Accession: T25893
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1173 <GAT>
A;Cross-references: EMBL:U80438; PIDN:AAB37645.1; GSPDB:GN00019; CESP:T19B4.7
A;Experimental source: strain Bristol N2; clone T19B4
C;Genetics:
A;Gene: CESP:T19B4.7
A;Map position: 1
A;Introns: 34/3; 85/1; 152/2; 455/2; 492/3; 499/1; 553/3; 673/2; 936/2; 1124/1

  Query Match              12.0%;  Score 156;  DB 2;  Length 1173;
  Best Local Similarity    26.4%;  Pred. No. 0.00041;
  Matches   57;  Conservative   38;  Mismatches   85;  Indels   36;  Gaps   10;

Qy       7 LLLVILFLPREMTSSVLTVNGKTENYILDTTP-------GSQASLICAVQNHTREEELLW 59
           ::| :  | |:  |  :     :   : |||       | :  | | | :  |  :: |
Db     191 IILKVTKLQRDFFSETV-----SNELVFITTPRLQVVEQGDEFLLECLVASLIR-PQVRW 244

Qy      60 YREEGR--VDLKSGNKINSSSVCVSSISENDNGISFTCRL-GRDQSVSVSVVLNVTFPPL 116
           ::  :  ||     ::  ||: || |  | | |: :|||    | |:  :| : |  ||
Db     245 IKDSPQIIVDGVPIRPVGVSSILVSPASIEDTGL-YTCRASNNDDSIDRAVSVEVRAPPP 303

Qy     117 LSGNDFQTVE-EGSNVKLVC-NVKANPQAQMMWYKN------SSLLDLEKSRHQIQQTSE 168
           ::      |  | ::|:| |    | |:|:: ||||      |    :| :|
Db     304 ITTRPTTKVAVETADVELECGTAAARPEARVNWYKNGEAIIGSEYFVIEPNR-------- 355

Qy     169 SFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIV 204
               | |  | : |   | |||:: : :|    |:|
Db     356 ---LRILGVVRADQAIYQCIAENDVGSEQASAQLLV 388


RESULT  12
T14316
rig 1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14316
S.Y.Ju, S.S.F., C.A., I.A., Dusiba, C.K., Lee, E.Y.H.P.
submitted to the EMBL Data Library, April 1998
A;Reference number: Z17975
A;Accession: T14316
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1344 <YUA>
A;Cross-references: EMBL:AF060570; NID:g4206385; PID:g4206386; PIDN:AAD11628.1

  Query Match              11.9%;  Score 153.5;  DB 2;  Length 1344;
  Best Local Similarity    29.7%;  Pred. No. 0.00075;
  Matches   68;  Conservative   38;  Mismatches   76;  Indels   47;  Gaps   16;

Qy       9 LVIL----FLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEG 64
           ||:|    || | :   ||                 :  : :| ||    :  | | :::|
Db     230 LVVLERPSFLRRPINQVVLA-------------DAPVNFLCEVQGDP-QPNLHWRKDDG 274

Qy      65 RVDLKSGN-KINSS-SVCVSSISENDNGISFTC----RLGRDQSVSVSVVLNVTFPP-LL 117
             :| :|  :| |  |: :  :|  | | ::||     :|| ::   |  |:|  ||  :
Db     275 --ELPAGRYEIRSDHSLWIDQVSSEDEG-TYTCVAENSVGRAEA---SGSLSVHVPPQFV 328

Qy     118 SGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQ-----ISESFQL 172
           :   |||  |:||   |  | ||   : | |  | : |  |:  :|       |  ||
Db     329 TKPQNQTVAPGANVSFQCETKGNPPPAIFWQKEGSQVLLFPSQ-SLQPMGRLLVSPRGQL 387

Qy     173 SITKVEKPDNGTYSCIAKS---SLKTESLDFHLIVKDKTV-GVPIEPII 217
           :|||:|  | | | | | |   |:  ::|   | :|  :: |:|  |||
Db     388 NITEVKIGDGGYYVCQAVSVAGSILAKALL---LEIKGASIDGLP--PII 431


RESULT  13
TDFFLK
protein tyrosine phosphatase (EC 3.1.3.48) DLAR precursor - fruit fly (Drosophila mel
N;Alternate names: leukocyte antigen-related protein
C;Species: Drosophila melanogaster
C;Date: 14-Dec-1990 #sequence_revision 02-May 1994 #text_change 22-Jun-1999
C;Accession: A36182
R;Streuli, M.; Krueger, N.X.; Tsai, A.Y.M.; Saito, H.
Proc. Natl. Acad. Sci. U.S.A. 86, 8698-8702, 1989
A;Title: A family of receptor linked protein tyrosine phosphatases in humans and Dros
A;Reference number: A36182; MUID:90046860
A;Accession: A36182
```

A;Molecule type: mRNA
A;Residues: 1-2029 <STR>
A;Cross-references: GB:M27700; NID:g157811; PIDN:AAA28668.1; PID:g157812
C;Genetics:
A;Gene: FlyBase:Lar
A;Cross-references: FlyBase:FBgn0000464
C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; ogy
C;Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-2029/Product: leukocyte antigen-related protein #status predicted <MAT>
F;33-1377/Domain: extracellular #status predicted <EXT>
F;50-113/Domain: immunoglobulin homology <IMM1>
F;154-211/Domain: immunoglobulin homology <IMM2>
F;249-303/Domain: immunoglobulin homology <IMM3>
F;321-401/Domain: fibronectin type III repeat homology <FN3A>
F;416-502/Domain: fibronectin type III repeat homology <FN3B>
F;514-599/Domain: fibronectin type III repeat homology <FN3C>
F;610-699/Domain: fibronectin type III repeat homology <FN3D>
F;708-802/Domain: fibronectin type III repeat homology <FN3E>
F;811-896/Domain: fibronectin type III repeat homology <FN3F>
F;909-993/Domain: fibronectin type III repeat homology <FN3G>
F;1006-1091/Domain: fibronectin type III repeat homology <FN3H>
F;1101-1198/Domain: fibronectin type III repeat homology <FN3I>
F;1378-1402/Domain: transmembrane #status predicted <TMM>
F;1403-2029/Domain: intracellular #status predicted <INT>
F;1417-2029/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F;1497-1718/Domain: protein-tyrosine-phosphatase homology <PTP1>
F;1786-2009/Domain: protein-tyrosine-phosphatase homology <PTP2>
F;57-111,161-209,256-301/Disulfide bonds: #status predicted
F;176/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1670/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1676/Binding site: substrate phosphate (Arg) #status predicted
F;1961/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1967/Binding site: substrate phosphate (Arg) #status predicted

Query Match 11.8%; Score 153; DB 1; Length 2029;
Best Local Similarity 25.9%; Pred. No. 0.0014;
Matches 51; Conservative 29; Mismatches 77; Indels 40; Gaps 9;

```
Qy      39 GSQASLICAVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISENDNGISF----T 94
           |  ||  || :      ::|        |:| |: | :    :: |   |||
Db      50 GGVASFYCAARGDP-PPSIVW--------RKNGKKV-SGTQSRYTVLEQPGGISILRIEP 99

Qy      95 CRLGRD------------QSVSVSVVLNV--------TFPPLLSGNDFQTVEEGSNVKL 133
           | |||              :||    | :         || :  |   : :| |  | :
Db     100 VRAGRDDAPYECVAENGVGDAVSADATLTIYEGDKTPAGFPVITQGPGTRVIEVGHTVLM 159

Qy     134 VCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSL 193
           |    ||   : | || : :|:   |: ::   : | | |    : | | | |:|::|:
Db     160 TCKAIGNPTPNIYWIKNQTKVDMSNPRYSLK---DGF-LQIENSPEEDQGKYECVAENSM 215

Qy     194 KTE-SLDFHLIVKDKTV 209
           || |   :| || : |
Db     216 GTEHSKATNLYVKVRRV 232
```

RESULT 14
T20992
hypothetical protein F15G9.4a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T20992; T24733
R;Sulston, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19355
A;Accession: T20992
A;Status: preliminary, translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5175 <WIL>
A;Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a
A;Experimental source: clone F15G9
R;Kershaw, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19929
A;Accession: T24733
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5175 <WIL>
A;Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a
A;Experimental source: clone T09B9
C;Genetics:
A;Gene: CESP:F15G9.4a
A;Map position: X
A;Introns: 85/1; 120/1; 444/3; 470/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/1; 2512/2; 2543/3; 2649/3; 2759/1; 2852/1; 2884/3; 2941/3; 2947/1; 2967/3; 2991/3; 4031/1; 4225/1; 4281/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 50

Query Match 11.8%; Score 152.5; DB 2; Length 5175;
Best Local Similarity 23.1%; Pred. No. 0.0045;
Matches 49; Conservative 42; Mismatches 98; Indels 23; Gaps 8;

```
Qy      33 ILDTTPGSQASLICAVQNHTR-EEELLWYREEGRVDLKSGNK-----INSSSVCVSSISE 86
           |:    |    | : :    : :::| |    :||:: :       |   : : :::|
Db    2199 IVTAIKGGALPFKCPIDDDKNFKGQIIWLRNYQPIDLEAEDARITRLSNDRRLTILNVTE 2258

Qy      87 NDNGISFTCRLGRDQSV-SVSVVLNVTFPP---LLSGNDFQTVEEGSNVKLVCNVKANPQ 142
           || |  ::||:  |     |      |  ||   :|  :  :|  | | | | |     |:
Db    2259 NDEG-QYSCRVKNDAGENSFDFKATVLVPPTIIMLDKDKNKTAVEHSTVTLSCPATGKPE 2317

Qy     143 AQMMWYKNSSLLDLEKSRHQIQQTS-ESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 201
             : |:|:   : :|     |        || ||:::: | | |:| | :|  :   | :
Db    2318 PDITWFKDGEAIHIENIADIIPNGELNGNQLKITRIKEGDAGKYTCEADNSAGSVEQDVN 2377

Qy     202 L----IVKDKTVGVPI-------EPIIAACVV 222
           :    | | :  |:|        | :: :| |
Db    2378 VNVITIPKIEKDGIPSDYESQQNERVVISCPV 2409
```

RESULT 15
T43290
hemicentin precursor - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C;Accession: T43290; T20993; T24734
R;Vogel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A;Description: Hemicentin is required for hemidesmosome mediated cell adhesion and ge
A;Reference number: Z22396
A;Accession: T43290
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5198 <VOG>
A;Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R;Sulston, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19355
A;Accession: T20993
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5198 <WIL>
A;Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b
A;Experimental source: clone F15G9
R;Kershaw, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19929
A;Accession: T24734
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5198 <WIL>
A;Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b
A;Experimental source: clone T09B9
C;Genetics:
A;Gene: him-4; F15G9.4b

A;Map position: X
A;Introns: 87..1, 1..1; 3344..3470, 1..1; 4772..4836, 1..1; 6641..6775, 1..1; 9772..10711, 1..1; 11843..
; 2..2, 2..2; 2644..2775, 1..1; 2884..2941, 1..1; 3487..3561, 1..1; 4033..
1; 4225..4361; 4408..4458; 4498..4547; 4848..4902; 4964..5034; 5100..

Query Match 11.8%; Score 152.5; DB 2; Length 5198;
Best Local Similarity 23.1%; Pred. No. 0.0046;
Matches 49; Conservative 42; Mismatches 98; Indels 23; Gaps 8;

```
Qy      33 ILDTTPGSQASLICAVQNHTR-EEELLWYREEGRVDLKSGNK-----INSSSVCVSSISE 86
           |:    |      | : :    : :::| |    :||:: :       |   : : :::|
Db    2199 IVTAIKGGALPFKCPIDDDKNFKGQIIWLRNYQPIDLEAEDARITRLSNDRRLTILNVTE 2258

Qy      87 NDNGISFTCRLGRDQSV-SVSVVLNVTFPP---LLSGNDFQTVEEGSNVKLVCNVKANPQ 142
           || |  ::||:  |    |       |  ||   :|  :  :|  | | | | |     |:
Db    2259 NDEG-QYSCKVKNDAGENSFDFKAIVLVPPTIMLDKDKNKIAVEHSTVILSCPATGKPE 2317

Qy     143 AQMMWYKNSSLIDLEKSRHQTQQTS-ESEQLSTIKVEKPENGTYSCIAKSSLKPESLDFR 201
           : |:|:   : :|     |         || ||:::: | | |:| | :|  :   | :
Db    2318 PDITWFKDGEAIHIENIADIIPNGELNGNQLKITRIKEGDAGKYTCEADNSAGSVEQDVN 2377

Qy     202 L----IVKDKTVCVPI-------EPIIAACVV 222
           :    | | :  |:|        | :: :| |
Db    2378 VNVITIPKIEKDGIPSDYESQQNERVVISCPV 2409
```

Search completed: September 17, 2002, 13:11:59
Job time: 11231 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:11:59 ; Search time 56.43 Seconds
(without alignments)
446.135 Million cell updates/sec

Title: US-09-863-823 6
Perfect score: 1336
Sequence: 1 MAWKSSVIMQMGRFLLLVIL.....RKKIMKICMKDKDPHSETAL 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	176	13.2	853	1	IJBONC	neural cell adhesi
2	175	13.1	858	1	IJRTNC	neural cell adhesi
3	173.5	13.0	2783	2	T34416	hypothetical prote
4	171.5	12.8	946	1	A47299	ror-related recept
5	170	12.7	1896	2	T08851	Down syndrome cell
6	166	12.4	761	1	IJHUNG	neural cell adhesi
7	161	12.1	1011	2	T13669	neuromusculin - fr
8	160.5	12.0	1091	1	IJCHNL	neural cell adhesi
9	159	11.9	725	1	IJMSNG	neural cell adhesi
10	159	11.9	1115	1	IJMSNL	neural cell adhesi
11	156	11.7	1173	2	T25893	hypothetical prote
12	153.5	11.5	1344	2	T14316	rig-1 protein - mo
13	153	11.5	2029	1	TDFFLK	protein-tyrosine-p
14	152.5	11.4	5175	2	T20992	hypothetical prote
15	152.5	11.4	5198	2	T43290	hemicentin precurs
16	150	11.2	1912	2	A56178	protein-tyrosine-p
17	149	11.2	1447	2	A54100	tumor suppressor p
18	148.5	11.1	338	2	JC4776	limbic-system-asso
19	148.5	11.1	4391	2	A38096	perlecan precursor
20	146.5	11.0	773	2	T46283	hypothetical prote
21	146	10.9	7962	2	I38346	elastic titin - hu
22	144.5	10.8	338	2	JC5519	50K glycoprotein p
23	144.5	10.8	1348	2	S51656	vascular endotheli
24	144	10.8	458	1	WMMSR1	biliary glycoprote
25	144	10.8	521	2	JC1508	biliary glycoprote
26	142.5	10.7	458	2	JC1509	biliary glycoprote
27	142.5	10.7	521	2	S34338	biliary glycoprote
28	142	10.6	871	1	I48696	protein-tyrosine k
29	142	10.6	881	1	I48697	protein-tyrosine k
30	142	10.6	1894	2	C54689	protein-tyrosine-p
31	141	10.6	976	2	T29583	hypothetical prote
32	139	10.4	1450	2	A44027	165K myofibrillar
33	138.5	10.4	797	2	S38579	fibroblast growth
34	138	10.3	1880	2	T18531	tractin - medicina
35	137.5	10.3	1898	2	S46216	leukocyte antigen-
36	137	10.3	725	2	JE0180	neural cell adhesi
37	137	10.3	1092	1	JN0635	neural cell adhesi
38	136.5	10.2	458	2	S23969	cell-adhesion mole
39	136.5	10.2	458	2	S68177	C CAM2a protein is
40	136.5	10.2	519	2	A44783	ecto-ATPase precur
41	136.5	10.2	1897	1	TDHULK	leukocyte antigen-
42	135	10.1	1273	2	T42405	sax-3 protein - Ca
43	135	10.1	1356	2	JC1402	protein-tyrosine k
44	134.5	10.1	1070	2	JC4593	protein-tyrosine k
45	134	10.0	1375	2	T13822	frazzled gene prot

ALIGNMENTS

RESULT 1
IJBONC
neural cell adhesion molecule short domain form precursor - bovine
N;Alternate names: NCAM-140
C;Species: Bos primigenius taurus (cattle)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 22-Jun-1999
C;Accession: A32976; A38778; B44290; S05402
R;Lipkin, V.M.; Khramtsov, N.V.; Andreeva, S.G.; Moshnyakov, M.V.; Petukhova, G.V.; R
FEBS Lett. 254, 69-73, 1989
A;Title: Calmodulin-independent bovine brain adenylate cyclase. Amino acid sequence a
A;Reference number: A32976; MUID:89378239
A;Accession: A32976
A;Molecule type: mRNA
A;Residues: 1-853 <LIP>
A;Cross-references: GB:X16451; NID:g60; PIDN:CAA34470.1; PID:g61
A;Accession: A38778
A;Molecule type: protein
A;Residues: 20-35;51-61;113-117;122-147;155-161;262-275;279-302;353-360;369-382;544-5
A;Note: the authors identified this protein as calmodulin-independent adenylate cycla
R;Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
A;Title: Structural and immunological characterization of the amino-terminal domain o
A;Reference number: A44290; MUID:86140120
A;Accession: B44290
A;Molecule type: protein
A;Residues: 20-36 <ROU>
A;Note: 23-Glu was also found
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM
C;Comment: Various forms of NCAM are produced by alternative splicing.
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; i
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-853/Product: neural cell adhesion molecule, short domain form #status experiment
F;20-719/Domain: extracellular #status predicted <EXT>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-288/Domain: immunoglobulin homology <IMM3>
F;261-270/Region: NCAM binding #status predicted
F;321-396/Domain: immunoglobulin homology <IMM4>
F;428-490/Domain: immunoglobulin homology <IMM5>
F;527-604/Domain: fibronectin type III repeat homology <FN3A>
F;633-693/Domain: fibronectin type III repeat homology <FN3B>
F;720-737/Domain: transmembrane #status predicted <TMM>
F;738-853/Domain: intracellular #status predicted <INT>
F;41-96,139-189,235-286,328-394,435-488/Disulfide bonds: #status predicted
F;222,314,346,432,458,487/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 13.2%, Score 176, DB 1, Length 853;
Best Local Similarity 28.4%; Pred. No. 1e-05;

```
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      49 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINS----SSVCVS 90
           | :| |                : :| : ||:  | | | :  |     : |    ||: :
Db     424 QVTLTCEASGDPIPSITWRTSTRNISSEEKASWTRPEKQETLDGHMVVRSHARVSSLTLK 483

Qy      91 SISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 146
           ||   | |  : |     :|:|   | |: | | : | | |       ||: | : | |
Db     484 SIQYTDAG-EYVCTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 538

Qy     147 ANPQAQMMWYKNSSLDDLEK-SRHQIQQTSESFQLSTTKVEKPDNGTYSCIAKSSLETES 205
           | | | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  ||
Db     539 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 598

Qy     206 LDFHLIVKDKTVGVP----IEPIIAACVVIF 232
           |:| |:  | |   |    :|| :   | |
Db     599 LEFVLVQAD-TPSSPSIDQVEPYSSTAQVQF 628


RESULT   2
I.RTNC
neural cell adhesion molecule short domain form precursor - rat
N;Alternate names: NCAM-140
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 22-Jun-1999
C;Accession: S00846; B37795; I58136
R;Small, S.J.; Shull, G.E.; Santoni, M.J.; Akeson, R.
J. Cell Biol. 105, 2335-2345, 1987
A;Title: Identification of a cDNA clone that contains the complete coding sequence for a
A;Reference number: S00846; MUID:88059265
A;Accession: S00846
A;Molecule type: mRNA
A;Residues: 1-858 <SMA>
A;Cross-references: EMBL:X06564
R;Small, S.J.; Akeson, R.
J. Cell Biol. 111, 2085-2096, 1990
A;Title: Expression of the unique NCAM VASE exon is independently regulated in distinct
A;Reference number: A37795; MUID:91035620
A;Accession: B37795
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 340-381 <SM2>
R;Small, S.J.; Haines, S.L.; Akeson, R.A.
Neuron 1, 1007-1017, 1988
A;Title: Polypeptide variation in an N-CAM extracellular immunoglobulin like fold is dev
A;Reference number: I58136; MUID:90166485
A;Accession: I58136
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 355-364 <RES>
A;Cross-references: GB:M32011; NID:g205643; PIDN:AAA41679.1; PID:g205644
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C;Comment: Various forms of NCAM are produced by alternative splicing.
C;Genetics:
A;Gene: NCAM
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; si
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-858/Product: neural cell adhesion molecule, short domain form #status predicted <MA
F;20-721/Domain: extracellular #status predicted <EXT>
F;44-98/Domain: immunoglobulin homology <IMM1>
F;142-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-290/Domain: immunoglobulin homology <IMM3>
F;263-272/Region: NCAM binding #status predicted
F;329-398/Domain: immunoglobulin homology <IMM4>
F;430-492/Domain: immunoglobulin homology <IMM5>
F;529-606/Domain: fibronectin type III repeat homology <FN3A>
F;645-695/Domain: fibronectin type III repeat homology <FN3B>
F;722-739/Domain: transmembrane #status predicted <TMM>
F;740-858/Domain: intracellular #status predicted <INT>
F;41-96,139-189,235-288,330-396,437-490/Disulfide bonds: #status predicted
F;222,316,348,434,462,489/Binding site: carbohydrate (Asn) (covalent) #status predict


  Query Match             13.1%;  Score 175;  DB 1;  Length 858;
  Best Local Similarity   28.4%;  Pred. No. 1.2e-05;
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      49 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINS----SSVCVS 90
           | :| |                : :| : ||:  | | | :  |     : |    ||: :
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEKASWIRPEKQETLDGHMVVRSHARVSSLTLK 384

Qy      91 SISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 146
           ||   | |  : |     :|:|   | |: | | : | | |       ||: | : | |
Db     385 SIQYTDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 440

Qy     147 ANPQAQMMWYKNSSLDDLEK-SRHQIQQTSESFQLSTTKVEKPDNGTYSCIAKSSLETES 205
           | | | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  ||
Db     441 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 500

Qy     206 LDFHLIVKDKTVGVP----IEPIIAACVVIF 232
           |:| |:  | |   |    :|| :   | |
Db     501 LEFILVQAD-TPSSPSIDRVEPYSSTAQVQF 530


RESULT   3
T34416
hypothetical protein F12F3.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34416
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F12F3.
A;Reference number: 221521
A;Accession: T34416
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2783 <FUL>
A;Cross-references: EMBL:AF000023; PIDN:AAC25886.1; GSPDB:GN00023; CESP:F12F3.2
A;Experimental source: strain Bristol N2; clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.2
A;Map position: 5
A;Introns: 45/3; 90/3; 471/3; 514/1; 2410/3; 2444/3; 2638/3; 2714/2; 2753/3; 2774/2;


  Query Match             13.0%;  Score 173.5;  DB 2;  Length 2783;
  Best Local Similarity   24.1%;  Pred. No. 6.5e-05;
  Matches   60;  Conservative   36;  Mismatches  100;  Indels   53;  Gaps    9;

Qy      12 GRFLLLVILPLPREMTSSVLTVNGKTENYILDTTP----------------------GS 48
           | |  |:   | :   |  :|:  |  :  | :||                      |
Db    1562 GIFNCLIENEIGQASASCQVIIFNKPAS--LQSTPDHSLEPNLVPTIQKALNNFSAQAGQ 1619

Qy      49 QASLICAVQNHTREEELLWYREE------GRVDLKSGNKINSSSVCVSSISENDNGISFT 102
           |  | | : : :  |  : |::::      || :| |  | |   ||  ::   | |  :
Db    1620 QIMLDPRISSPS-ESTVAWFKDDERIFSAGRYELSSDKKSNHKLVC-HAVQSQDTG-KYR 1676

Qy     103 C----RLGRDQS---VSVSVVLNVTFPPLLSGNDFQTVEFGSNVKLVCNVKANPQAQMMW 155
           |    : |  :|   |:|  |     |  :     |   | |: | | | :  : |  :|
Db    1677 CVVINPYGYAESFCNVAVEDVIKFIAPSFSATISDSTAIIGHNITLECKVFGSPAPPVSW 1736

Qy     156 YKNSSLIDLEKSRHQIQQTSFSFQLSTTKVEKPDNGTYSCIAKS---------  SLKT 203
           |:   :   :   | |  : :|||:| |  | | | | | |             ||
Db    1737 TKDGFRISTTFRIRQTQPENGNCKISSKAFSDDMGVYVCSATSVAGVDSTSSMVMIAKT 1796

Qy     204 ESLDFHLIV 212
             | ||::
Db    1797 TCTDSHLVI 1805


RESULT   4
A47299
```

```
ror-related receptor RTK - Pacific electric ray
N;Contains: protein-tyrosine kinase (EC 2.7.1.112)
C;Species: Torpedo californica (Pacific electric ray)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-May-2000
C;Accession: A47299
R;Jennings, C.G.; Dyer, S.M.; Burden, S.J.
Proc. Natl. Acad. Sci. U.S.A. 90, 2895-2899, 1993
A;Title: Muscle-specific trk-related receptor with a kringle domain defines a distinct c
A;Reference number: A47299; MUID:93219391
A;Accession: A47299
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-946 <JEN>
A;Cross-references: GB:L11311; NID:g290857; PIDN:AAA49285.1; PID:g290858
A;Experimental source: electric organ
A;Note: sequence extracted from NCBI backbone (NCBIN.128724, NCBIP.128726)
C;Superfamily: Torpedo ror-related receptor; immunoglobulin homology; kringle homology;
C;Keywords: ATP; glycoprotein; kringle; phosphotransferase; transmembrane protein; tyros
F;44-103/Domain: immunoglobulin homology <IM1>
F;137-195/Domain: immunoglobulin homology <IM2>
F;229-287/Domain: immunoglobulin homology <IM3>
F;464-542/Domain: kringle homology <KR3>
F;572-588/Domain: transmembrane #status predicted <TMN>
F;650-940/Domain: protein kinase homology <KIN>
F;658-666/Region: protein kinase ATP-binding motif
F;225,340,477,544/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             12.8%;  Score 171.5;  DB 1;  Length 946;
  Best Local Similarity   25.1%;  Pred. No. 2.5e-05;
  Matches   63;  Conservative   38;  Mismatches   95;  Indels   55;  Gaps   10;

Qy      15 LLLVIL---------FLPR--EMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREE 63
           ||:: |          ||:  ::|| : ||:   |          :|| :||| ::
Db      10 LLMIFLVTTGGSADGILPKAPQITSPLETVDALVE---------EEASFMCAVDSYP-AA 59

Qy      64 ELLWYREEGRV---DLKSGNKINSSSVCVSSISENDNGI-SFTCRLGRDQSVSVSVVLNV 119
           |: | |    :   | :  | |   : :|: ||| :   |   |   |      | |
Db      60 EITWTRNNIPIRPFDTRYSTKENGQILTILSVEDTDNGVYCCTANNGMGSSAQSCGALQV 119

Qy     120 TFPP--LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSES 177
             |  :    | : :  || | | |:   ||:  : |:|: : |  :: |  :    ||
Db     120 KMKPKIIRPPTDVRAL-LGSKVVLPCSTMGNPKPAISWFKDETALKNDQPRTSV---LES 175

Qy     178 FQLSITKVEKPDNGTYSCIAKSSLKTE----------------------SLDFHLIVK 213
           | |  |:  | | | |:|:::||  |                      |   :|::
Db     176 GNLRIRNVQLEDAGKYRCLARNSLGFEYSRSAALEVQVSARIVKAPTSQNVSYGSEVILQ 235

Qy     214 DKTVGVPIEPI 224
           |  | ||  |
Db     236 CKATGFPIPTI 246


RESULT   5
T08851
Down syndrome cell adhesion protein 1 - human (fragment)
N;Alternate names: Down syndrome cell adhesion molecule
C;Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08851
R;Yamakawa, K.; Huo, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.E.; Korenberg
submitted to the EMBL Data Library, September 1997
A;Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down sy
A;Reference number: Z16495
A;Accession: T08851
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1896 <YAM>
A;Cross-references: EMBL:AF023449; NID:g3169765; PID:g3169766
A;Experimental source: brain; developmental stage: 14 weeks; fetal
C;Genetics:
A;Gene: DSCAM
A;Map position: 21q22
A;Note: derived from alternately spliced mRNA
C;Function:
A;Description: involved in nervous system development
C;Keywords: alternative splicing


  Query Match             12.7%;  Score 170;  DB 2;  Length 1896;
  Best Local Similarity   29.8%;  Pred. No. 7.6e-05;
  Matches   57;  Conservative   31;  Mismatches   69;  Indels   34;  Gaps   10;

Qy      23 PREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKSGNK- 81
           ||:: |||                 ||| || |:|   | ::|| ||| | : |  |
Db     306 PRKVKSSV---------------GSQVSLSCSVTG-TEDQELSWYR-NGEI-LNPGKNV 346

Qy      82 ----INSSSVCVSSISENDNGISFTCRLGRDQSVS---VSVVLNVTFPPLLSGNDFQTVE 134
               ||  .. .  . ..| | .. | . .|:  :   | |||    | ::|    : |
Db     347 RITGINHENLIMDHMVKSDGG-AYQCFVRKDKLSAQDYVQVVLEDGTPKIISAFSEKVVS 405

Qy     135 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQ--TSES---FQLSITKVEKPD 189
               | |:|||  |   : | :   : |:   |:| |  |||      |:|:  :  |
Db     406 PAEPVSLMCNVKGIPLPTITWILDDPI-LKGGSHKISQMIISEGNVVSYLNISSSQVRD 464

Qy     190 NGTYSCIAKSS 200
           | | | | :|
Db     465 GGVYRCTANNS 475


RESULT   6
IJHUNG
neural cell adhesion molecule 1 GPI anchored splice form precursor, muscle-specific -
N;Alternate names: CD56; NCAM-120
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 28-Jan-2000
C;Accession: S07784; A26883
R;Barton, C.H.; Dickson, G.; Gower, H.J.; Rowett, L.H.; Putt, W.; Elsom, V.; Moore, S
Development 104, 165-173, 1988
A;Title: Complete sequence and in vitro expression of a tissue-specific phosphatidyli
A;Reference number: S07784; MUID:89305258
A;Accession: S07784
A;Molecule type: mRNA
A;Residues: 1-761 <BAR>
A;Cross-references: EMBL:X16841; NID:g35005; PIDN:CAA34739.1; PID:g35006
R;Dickson, G.; Gower, H.J.; Barton, C.H.; Prentice, H.M.; Elsom, V.L.; Moore, S.E.; C
Cell 50, 1119-1130, 1987
A;Title: Human muscle neural cell adhesion molecule (N CAM): identification of a musc
A;Reference number: A90895; MUID:87301755
A;Accession: A26883
A;Molecule type: mRNA
A;Residues: 491-761 <DIC>
A;Cross-references: GB:M17409; NID:g189097; PIDN:AAA59912.1; PID:g386979
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM
C;Comment: Various forms of NCAM are produced by alternative splicing.
C;Genetics:
A;Gene: GDB:NCAM1; NCAM; CD56
A;Cross-references: GDB:119448; OMIM:116930
A;Map position: 11q22.2-11q22.3
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; i
C;Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membra
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-761/Product: neural cell adhesion molecule phosphatidylinositol-linked form, mus
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-289/Domain: immunoglobulin homology <IMM3>
F;263-272/Region: NCAM binding #status predicted
F;322-387/Domain: immunoglobulin homology <IMM4>
F;419-481/Domain: immunoglobulin homology <IMM5>
F;499-587/Domain: fibronectin type III repeat homology #status atypical <FN3A>
F;633-720/Domain: fibronectin type III repeat homology #status atypical <FN3B>
F;41-96,139-189,235-287,329-385,426-479/Disulfide bonds: #status predicted
F;222,315,347,423,449,478/Binding site: carbohydrate (Asn) (covalent) #status predict
```

```
  Query Match             12.4%;  Score 166;  DB 1;  Length 761;
  Best Local Similarity   28.0%;  Pred. No. 5.2e-05;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      49 QASLIC ------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 94
           |  :| |             : :|  : ||: |    :|  : ::|  ::  ||: : ||
Db     424 QVTLTCEASGDPIPSITWRTSTRNISSEEKTL----DGHMVVRSHARV--SSLTLKSIQY 477

Qy      95 NENGSFTC    RTGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVFLVCNVKANPQ 150
           | |  : |     :|:|   | |: | | : | | |       ||: | : | | | |
Db     478 TDAG-EYICIASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 533

Qy     151 AQMMWYKNSSLLDLEK-SPHQTQQTSFSEQLSLTKVEKPDNGTYSCIAKSSLETESLDEH 209
           | : |:::  ||     |  :| |  :  | :|   : | | |:| | : :  |||:|
Db     534 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 593

Qy     210 LIVKDKTVGVP----IEPITAACVVIF 232
           |:  | |   |    :||  :   | |
Db     594 LVQAD-TPSSPSIDQVEPYSSTAQVQF 619


RESULT   7
T13669
neuromusculin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13669
R;Kania, A.; Han, P.L.; Kim, Y.T.; Bellen, H.
Neuron 11, 673-687, 1993
A;Title: Neuromusculin, a Drosophila gene expressed in peripheral neuronal precursors an
A;Reference number: Z17697; MUID:94000831
A;Accession: T13669
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1011 <KAN>
A;Cross-references: EMBL:L23146; NID:g385073; PID:g385074; PIDN:AAA03750.1
C;Genetics:
A;Gene: nrm
A;Cross-references: FlyBase:FBgn0005623


  Query Match             12.1%;  Score 161;  DB 2;  Length 1011;
  Best Local Similarity   24.8%;  Pred. No. 0.00018;
  Matches   46;  Conservative   34;  Mismatches   81;  Indels   28;  Gaps    6;

Qy      41 LTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-----EEGRVDL--------- 76
           : :||   :    |  ||:  | | :          | ||        |  ::
Db     265 IDINGVKHH----TVQGSKVVLTCDTHGARPAVNLTWYNTTTTISSGENETTEVRSKSLE 320

Qy      77 KSGNKINSSSVCVSSISFNENGTSFTC-------RTGRDQSVSVSVVLNVTFPPLLSGND 129
           ||    :: |  : : :  :|   | |          :: |:: :| :: | | :|::  :
Db     321 KSDGTFHTQSELIFNATRFENDRVFRCEAENIVLQINREKPISSALTLEVLYPPVVKVSP 380

Qy     130 FQTVEEGSNVKLV-CNVKANPQA--QMMWYKNSSLLDLEKSRHQTQQTSESFQLSLTKVE 186
                 | : |: |   ||| :  |: ||:|  |::   : |     ||:  | |   |
Db     381 SAITANTSEIVLLNCEYFANPASLTQVEWYRNDILVNVNDTTHYKCGGNSFNVALVIKSTE 440

Qy     187 KPDNGTYSC 195
           | | | |||
Db     441 KEDIGNYSC 449


RESULT   8
IJCHNL
neural cell adhesion molecule long domain form precursor - chicken
N;Alternate names: NCAM-180
N;Contains: neural cell adhesion molecule, short domain form (NCAM-140)
C;Species: Gallus gallus (chicken)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 22-Jun-1999
C;Accession: A43613; B43613; A25435; B25435; A46550; S36950; A44369; A60852; S29668
R;Cunningham, B.A.; Hemperly, J.J.; Murray, B.A.; Prediger, E.A.; Brackenbury, R.; Edelm
Science 236, 799-806, 1987

A;Title: Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell
A;Reference number: A43613; MUID:87206190
A;Accession: A43613
A;Molecule type: mRNA
A;Residues: 1-175 <CU2>
A;Cross-references: GB:M15860
A;Accession: B43613
A;Molecule type: protein
A;Residues: 20-44,120-127,202-221,320-342,399-415,640-659,822-828 <CU3>
A;Note: Asn-222 probably binds carbohydrate; Asn-226 probably does not
R;Hemperly, J.J.; Murray, B.A.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 3037-3041, 1986
A;Title: Sequence of a cDNA clone encoding the polysialic acid-rich and cytoplasmic d
A;Reference number: A25435; MUID:86206089
A;Accession: A25435
A;Molecule type: mRNA
A;Residues: 128-1091 <HEM>
A;Cross-references: GB:M13210
A;Accession: B25435
A;Molecule type: protein
A;Residues: 128-149;222-249;428-459;511-561;744-769;784-781;1080-1084 <HE2>
R;Murray, B.A.; Owens, G.C.; Prediger, E.A.; Crossin, K.L.; Cunningham, B.A.; Edelman
J. Cell Biol. 103, 1431-1439, 1986
A;Title: Cell surface modulation of the neural cell adhesion molecule resulting from
A;Reference number: A46550; MUID:87033934
A;Accession: A46550
A;Molecule type: DNA
A;Residues: 810-1070 <MUR>
A;Cross-references: GB:X04479
R;Sasner, M.; Covault, J.
submitted to the EMBL Data Library, February 1993
A;Reference number: S36950
A;Accession: S36950
A;Molecule type: DNA
A;Residues: 1-17 <SAS>
A;Cross-references: EMBL:X70342; NID:g417631; PIDN:CAA49807.1; PID:g417632
R;Caldwell, G.; Li, R.; Forrest, D.; Brackenbury, R.
Genomics 14, 875-882, 1992
A;Title: Conserved regulatory elements in the promoter region of the N-CAM gene.
A;Reference number: A44369; MUID:93122797
A;Accession: A44369
A;Molecule type: DNA
A;Residues: 1-17 <COE>
A;Cross-references: EMBL:Z12128; NID:g63653; PIDN:CAA78113.1; PID:g63654
A;Experimental source: White Leghorn
R;Cole, G.J.; Loewy, A.; Cross, N.V.; Akeson, R.; Glaser, L.
J. Cell Biol. 103, 1739-1744, 1986
A;Title: Topographic localization of the heparin-binding domain of the neural cell ad
A;Reference number: A60852; MUID:87057627
A;Accession: A60852
A;Molecule type: protein
A;Residues: 20-29 <COL>
R;Rao, Y.; Wu, X.F.; Gariepy, J.; Rutishauser, U.; Siu, C.H.
J. Cell Biol. 118, 937-949, 1992
A;Title: Identification of a peptide sequence involved in homophilic binding in the n
A;Reference number: A43280; MUID:92363934
A;Contents: annotation; homophilic binding region
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM
C;Comment: Various forms of NCAM are produced by alternative splicing.
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; i
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1091/Product: neural cell adhesion molecule, long domain form #status experiment
F;20-809,1071-1091/Product: neural cell adhesion molecule, short domain form #status
F;20-711/Domain: extracellular #status predicted <EXT>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-284/Domain: immunoglobulin homology <IMM3>
F;262-271/Region: NCAM binding #status experimental
F;322-387/Domain: immunoglobulin homology <IMM4>
F;419-481/Domain: immunoglobulin homology <IMM5>
```

```
F;518-595/Domain: fibronectin type III repeat homology <FN3A>
F;624-685/Domain: fibronectin type III repeat homology <FN3B>
F;712-729/Domain: transmembrane #status predicted <TMM>
F;730-1091/Domain: intracellular #status predicted <INT>
F;41-96,139-189,235-287,329-385,426-479/Disulfide bonds: #status predicted
F;222/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;315,347,423,449,478/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             12.0%;  Score 160.5;  DB 1;  Length 1091;
  Best Local Similarity   26.9%;  Pred. No. 0.00021;
  Matches   60;  Conservative   37;  Mismatches   87;  Indels   39;  Gaps   10;

Qy      33 VNGKTENYILDTTPGSQASLIC--------------AVQNHTREEELLWYREEGRVDLKS 78
           |  ||   : |     | :| |               : :| : ||: |    :||: ::|
Db     313 VENKTAMELED-----QITLTCEASGDPIPSITWKTSTRNISNEEKTL----DGRIVVRS 363

Qy      79 GNKINSSSVCVSSISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVE 134
             ::  ||: :  |   | |  : |     :|:|   | :: | | : | | |
Db     364 HARV--SSLTLKEIQYTDAG-EYVCTASNTIGQD---SQAMYLEVQYAPKLQGPVAVYTW 417

Qy     135 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTY 193
           ||: | : | | | | | | : |::  ||     |  :|  |  :  | :|   : | | |
Db     418 EGNQVNITCEVFAYPSAVISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNY 477

Qy     194 SCIAKSSLKTESLDFHLIVKDKTVGVP----IEPIIAACVVIF 232
           :| | :  || :| |:  | |   |     :||  :   | |
Db     478 NCTAVNRIGQESSEFILVQAD-TPSSPSIDRVEPYSSTARVEF 519


RESULT   9
IJMSNG
neural cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse
N;Alternate names: NCAM-120
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 31-Dec-2000
C;Accession: A29673; S00382; A44290
R;Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fontec
EMBO J. 6, 907-914, 1987
A;Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000
A;Reference number: A29673; MUID:87246524
A;Accession: A29673
A;Molecule type: mRNA
A;Residues: 1-725 <BAR>
A;Cross-references: EMBL:Y00051; NID:g53342; PIDN:CAA68263.1; PID:g53343
R;Barbas, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.
EMBO J. 7, 625-632, 1988
A;Title: Differential splicing and alternative polyadenylation generates distinct NCAM t
A;Reference number: S00382; MUID:88283628
A;Accession: S00382
A;Molecule type: DNA
A;Residues: 642-656,'D',658-725 <BA2>
A;Cross-references: EMBL:X07195
R;Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
A;Title: Structural and immunological characterization of the amino-terminal domain of m
A;Reference number: A44290; MUID:86140120
A;Accession: A44290
A;Molecule type: protein
A;Residues: 20-36 <ROU>
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C;Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:IJMS
C;Genetics:
A;Gene: NCAM
A;Map position: 9
A;Introns: 701/1
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane
F;1-19/Domain: signal sequence #status predicted <SIG>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-296/Domain: immunoglobulin homology <IMM3>
F;263-272/Region: NCAM binding #status predicted
F;323-388/Domain: immunoglobulin homology <IMM4>
F;420-482/Domain: immunoglobulin homology <IMM5>
F;519-596/Domain: fibronectin type III repeat homology <FN3A>
F;625-685/Domain: fibronectin type III repeat homology <FN3B>
F;41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted
F;222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predict


  Query Match             11.9%;  Score 159;  DB 1;  Length 725;
  Best Local Similarity   27.5%;  Pred. No. 0.00017;
  Matches   57;  Conservative   35;  Mismatches   81;  Indels   34;  Gaps    9;

Qy      49 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 94
           | :| |               : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEQDL----DGHMVVRSHARV--SSLTLKSIQY 378

Qy      95 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 150
             | |  : |      :|:|   | ||: |   : | | |       ||: | : | | | |
Db     379 RDAG-EYMCTASNTIGQD---SQSIDLEFQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     151 AQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 209
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     210 LIVKDKTVGVP----IEPIIAACVVIF 232
           |:  | |   |    :||  :   | |
Db     495 LVQAD-TPSSPSIDRVEPYSSTAQVQF 520


RESULT  10
IJMSNL
neural cell adhesion molecule 1 precursor, long domain splice form - mouse
N;Alternate names: NCAM-180
N;Contains: neural cell adhesion molecule, short domain splice form (NCAM-140)
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 31-Dec-2000
C;Accession: A29673; S00844; S00384; A28281; A44290; S00383
R;Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fon
EMBO J. 6, 907-914, 1987
A;Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,
A;Reference number: A29673; MUID:87246524
A;Accession: A29673
A;Molecule type: mRNA
A;Residues: 1-548,'T',550-571,'T',573-574,'D',576-588,'MQPS',593,'S',595-599,'P',601,
A;Cross-references: EMBL:Y00051; NID:g53342; PIDN:CAA68263.1; PID:g53343
R;Santoni, M.J.; Barthels, D.; Barbas, J.A.; Hirsch, M.R.; Steinmetz, M.; Goridis, C.
Nucleic Acids Res. 15, 8621-8641, 1987
A;Title: Analysis of cDNA clones that code for the transmembrane forms of the mouse n
A;Reference number: S00844; MUID:88067687
A;Accession: S00844
A;Molecule type: mRNA
A;Residues: 529-809,1077-1115 <SAN>
A;Cross-references: EMBL:X06328; NID:g53322; PIDN:CAA29641.1; PID:g817984
R;Barbas, J.A., Chaix, J.C.; Steinmetz, M.; Goridis, C.
EMBO J. 7, 625-632, 1988
A;Title: Differential splicing and alternative polyadenylation generates distinct NCA
A;Reference number: S00382; MUID:88283628
A;Accession: S00384
A;Molecule type: DNA
A;Residues: 642-1115 <BAR>
A;Cross-references: EMBL:X07195
R;Barthels, D.; Vopper, G.; Wille, W.
Nucleic Acids Res. 16, 4217-4225, 1988
A;Title: NCAM-180, the large isoform of the neural cell adhesion molecule of the mous
A;Reference number: A28281; MUID:88247737
A;Accession: A28281
A;Molecule type: mRNA
A;Residues: 804-1081 <BA3>
A;Cross-references: EMBL:X07244; NID:g53321; PIDN:CAA30230.1; PID:g929720
R;Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
```

```
A;Title: Structural and immunological characterization of the amino-terminal domain of
A;Reference number: A44290; MUID:86140120
A;Accession: A44290
A;Molecule type: protein
A;Residues: 20-36 <ROD>
C;Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C;Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:IJMS
C;Genetics:
A;Gene: NCAM
A;Map position: 9
A;Introns: 643/3; 701/1; 770/2; 809/2; 1076/2
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; si
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1115/Product: neural cell adhesion molecule, long domain splice form #status experi
F;20-809,1077-1115/Product: neural cell adhesion molecule, short domain splice form #sta
F;20-711/Domain: extracellular #status predicted <EXT>
F;34-98/Domain: immunoglobulin homology <IMM1>
F;132-191/Domain: immunoglobulin homology <IMM2>
F;152-156/Region: heparin binding #status predicted
F;161-165/Region: heparin binding #status predicted
F;228-290/Domain: immunoglobulin homology <IMM3>
F;262-272/Region: NCAM binding #status predicted
F;323-388/Domain: immunoglobulin homology <IMM4>
F;420-482/Domain: immunoglobulin homology <IMM5>
F;519-596/Domain: fibronectin type III repeat homology <FN3A>
F;625-685/Domain: fibronectin type III repeat homology <FN3B>
F;712-729/Domain: transmembrane #status predicted <TMM>
F;730-1115/Domain: intracellular #status predicted <INT>
F;41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted
F;222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match              11.9%;  Score 159;  DB 1;  Length 1115;
  Best Local Similarity    27.5%;  Pred. No. 0.00028;
  Matches   57;  Conservative   35;  Mismatches   81;  Indels   34;  Gaps    9;

Qy      49 QASLCD              AVQNHTREEELLWYRFFG 84
           | :| |              : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEQDL----DGHMVVRSHARV--SSLTLKSIQY 378

Qy      85 NDNGISFTC ---RLGRDQSVSVSVVLNVTFPPLLSGNDFQIVEEGSNVKLVCNVKANPQ 140
             | |  : |     :||:|   | |: |   : | | |       ||: | : | | | |
Db     379 RDAG-EYMCTASNTIGQD---SQSIDLEFQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     141 AQMMWYKNSSLLDLEK-SRHQTQQISESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 209
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     210 LIVKDKTVGVP----IEPIIAACVVIF 232
           |:  | |  |     :||  :  | |
Db     495 LVQAD-TPSSPSIDRVEPYSSTAQVQF 520


RESULT  11
T25893
hypothetical protein T19B4.7  Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25893
R;Gattung, S.
submitted to the EMBL Data Library, November 1996
A;Description: The sequence of C. elegans cosmid T19B4.
A;Reference number: Z20196
A;Accession: T25893
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1173 <GAT>
A;Cross-references: EMBL:U80438, PIDN:AAB37635.1; CESP:T19B4.7
A;Experimental source: strain Bristol N2; clone T19B4
C;Genetics:
A;Gene: CESP:T19B4.7
A;Map position: 1

A;Introns: 24/3; 96/1; 154/2; 442/2; 499/1; 55323; 57322; 57465; 11.4/1

  Query Match              11.7%;  Score 156;  DB 2;  Length 1173;
  Best Local Similarity    26.4%;  Pred. No. 0.00051;
  Matches   57;  Conservative   38;  Mismatches   85;  Indels   36;  Gaps   10;

Qy      15 LLLVILFLPREMTSSVLTVNGKTENYILDTTP-------GSQASLICAVQNHTREEELLW 67
           ::| :  | |: |  :     :  : |||       | :  | | | :  |  :: |
Db     191 IILKVTKLQRDFFSETV-----SNELVFTTPRLQVVEQGDEFLLECLVASLIR-PQVRW 244

Qy      68 YREEGS  VDLKSGNKINSSSVCVSSISENDNGISFTCRL GRDQSVSVSVVLNVTFPPL 124
            ::  :  ||     ::  ||: ||  |  | |: :|||    | |:  :| : |  ||
Db     245 LKDSRQIIVLGVPIRRVGVSSILVSKASIEDRGL-YTCRASNNDDSIDRAVSVEVRAPPR 303

Qy     125 ISGNDFQIVE-EGSNVKLVC-NVKANPQAQMMWYKN------SSLLDLEKSRHQTQQISF 176
           ::      | | ::|:| |    | |:|:: ||||      |    :| :|
Db     304 ITTRPTTKVAVETADVELECGTAAARPEARVNWYKNGEAIIGSEYFVIEPNR-------- 355

Qy     177 SFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIV 212
              | |  | : |   | |||:: : :|     |:|
Db     356 ---LRILGVVRADQAIYQCIAENDVGSEQASAQLLV 388


RESULT  12
T14316
rig-1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 26-Sep-1999 #sequence_revision 26-Sep-1999 #text_change 26-Sep-1999
C;Accession: T14316
R;Yuan, S.C.F., Cox, L.A., Dasika, G.K.; Lee, E.Y.H.P.
submitted to the EMBL Data Library, April 1998
A;Reference number: Z17975
A;Accession: T14316
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1344 <YUA>
A;Cross-references: EMBL:AF040975; NID:g4206485; PIDN:AAC31628.1


  Query Match              11.5%;  Score 153.5;  DB 2;  Length 1344;
  Best Local Similarity    29.7%;  Pred. No. 0.00095;
  Matches   68;  Conservative   38;  Mismatches   76;  Indels   47;  Gaps   16;

Qy      17 LVIL----FLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEFLLWYRFFG 72
           ||:|    || | :   ||                :  : :| ||    :  | | :::|
Db     230 LVVLERPSFLRRPINQVVLA--------------DAPVNFLCEVQGDP-QPNLHWRKDDG 274

Qy      73 RVDLKSGN-KINSS-SVCVSSISENDNGISFTC----RLGRDQSVSVSVVLNVTFPP-LL 125
             :| :|  :| |  |: :  :|  | | ::||     :|| ::   |  |:|  ||  :
Db     275 --KLPAGRYEIRSDHSLWIDQVSSEDEG-IYICVAENSVGKAEA---SGSLSVHVPPQFV 328

Qy     126 SGNDFQIVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQTQQ------ISESFQL 180
           :    ||| |:||   |  | ||   : | |  | : |  |:  :|       |   ||
Db     329 TKPQNQTVAPGANVSFQCETKGNPPPAIFWQKEGSQVLLFPSQ-SLQPMGRLLVSPRGQL 387

Qy     181 SITKVEKPDNGTYSCIAKS---SLKTESLDFHLIVKDKTV-GVPIEPII 225
           :||:|:  | | | | | |   |:  ::|   | :|  :: |:|  |||
Db     388 NITEVKIGDGGYYVCQAVSVAGSILAKAL---LEIKGASIDGLP--PII 431


RESULT  13
TDFFLK
protein-tyrosine phosphatase (EC 3.1.3.48) DLAR precursor  fruit fly (Drosophila mel
N;Alternate names: leukocyte antigen related protein
C;Species: Drosophila melanogaster
C;Date: 14-Dec-1990 #sequence_revision 02-May-1994 #text_change 22-Jun-1999
C;Accession: A36182
R;Streuli, M.; Krueger, N.X.; Tsai, A.Y.M.; Saito, H.
Proc. Natl. Acad. Sci. U.S.A. 86, 8698-8702, 1989
A;Title: A family of receptor-linked protein tyrosine phosphatases in humans and Dros
A;Reference number: A36182; MUID:90046860
A;Accession: A36182
```

A;Molecule type: mRNA
A;Residues: 1-2029 <STR>
A;Cross-references: GB:M27700, NID.g157811, PIDN.AAA28668.1, PID.g157812
C;Genetics:
A;Gene: FlyBase:Lar
A;Cross-references: FlyBase:FBgn0000464
C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; ogy
C;Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-2029/Product: leukocyte antigen-related protein #status predicted <MAT>
F;33-1377/Domain: extracellular #status predicted <EXT>
F;50-113/Domain: immunoglobulin homology <IMM1>
F;154-211/Domain: immunoglobulin homology <IMM2>
F;249-303/Domain: immunoglobulin homology <IMM3>
F;321-401/Domain: fibronectin type III repeat homology <FN3A>
F;416-502/Domain: fibronectin type III repeat homology <FN3B>
F;514-599/Domain: fibronectin type III repeat homology <FN3C>
F;610-699/Domain: fibronectin type III repeat homology <FN3D>
F;708-802/Domain: fibronectin type III repeat homology <FN3E>
F;811-896/Domain: fibronectin type III repeat homology <FN3F>
F;909-993/Domain: fibronectin type III repeat homology <FN3G>
F;1006-1091/Domain: fibronectin type III repeat homology <FN3H>
F;1101-1198/Domain: fibronectin type III repeat homology <FN3I>
F;1378-1402/Domain: transmembrane #status predicted <TMM>
F;1403-2029/Domain: intracellular #status predicted <INT>
F;1417-2029/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F;1497-1718/Domain: protein-tyrosine phosphatase homology <PTP1>
F;1786-2009/Domain: protein-tyrosine-phosphatase homology <PTP2>
F;57-111,161-209,256-301/Disulfide bonds: #status predicted
F;176/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1670/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1676/Binding site: substrate phosphate (Arg) #status predicted
F;1961/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1967/Binding site: substrate phosphate (Arg) #status predicted

Query Match 11.5%; Score 153; DB 1; Length 2029;
Best Local Similarity 25.9%; Pred. No. 0.0017;
Matches 51; Conservative 29; Mismatches 77; Indels 40; Gaps 9;

```
Qy      47 GSQASLICAVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISENDNGISF----T 102
           |  ||  || :       ::|         |:| |: | :    :: |   |||
Db      50 GGVASFYCAARGDP-PPSIVW--------RKNGKKV-SGTQSRYTVLEQPGGISILRIEP 99

Qy     103 CRLGRD------------QSVSVSVVLNV--------TFPPLLSGNDFQTVEEGSNVKL 141
           | |||              :||    | :         || :  |   : :| |  | :
Db     100 VRAGRDDAPYECVAENGVGDAVSADATLTIYEGDKTPAGFPVITQGPGTRVIEVGHTVLM 159

Qy     142 VCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSL 201
           |    ||   : | || : :|:   |: ::   : | | |    : | | | |:|::|:
Db     160 TCKAIGNPTPNIYWIKNQTKVDMSNPRYSLK---DGF-LQIENSREEDQGKYECVAENSM 215

Qy     202 KTE-SLDFHLIVKDKTV 217
           || |   :| || : |
Db     216 GTEHSKATNLYVKVRRV 232
```

RESULT 14
T20992
hypothetical protein F15G9.4a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T20992; T24733
R;Sulston, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19355
A;Accession: T20992
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5175 <WIL>
A;Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a
A;Experimental source: clone F15G9
R;Kershaw, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19929
A;Accession: T24733
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5175 <WI2>
A;Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a
A;Experimental source: clone T09B9
C;Genetics:
A;Gene: CESP:F15G9.4a
A;Map position: X
A;Introns: 85/1; 123/2; 334/3; 470/1; 477/2; 598/1; 664/1; 945/1; 997/1; 1051/3; 1184/1; 2512/2; 2544/3; 2644/1; 2759/1; 2812/1; 2889/1; 2913/3; 2941/1; 2967/2; 2991/3; 3031/3; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4547/2; 4838/1; 4879/1; 4941/1; 5011/1; 5

Query Match 11.4%; Score 152.5; DB 2; Length 5175;
Best Local Similarity 23.1%; Pred. No. 0.0057;
Matches 49; Conservative 42; Mismatches 98; Indels 23; Gaps 8;

```
Qy      41 ILDTTPGSQASLICAVQNHTR-EEELLWYREEGRVDLKSGNK-----INSSSVCVSSISE 94
           |:    |      | : :    : :::| |    :||:: :      |   : : :::|
Db    2199 IVIAIKGGALPIKPPIDDDKNPKGQIIWLPNYQPIPLEAEDARITPLSNDPPPITIINVTF 2258

Qy      95 NDNGISFTCRLGRDQSV-SVSVVLNVTFPP---LLSGNDFQTVEEGSNVKLVCNVKANPQ 150
           || |  ::||:  |    |       |  ||   :|  :  :|  | | | | |     |:
Db    2259 NDEC QYSCRVKNDAGENSFDFKATVLVPPTIIMLDKDKNKTAVEHSTVTLSCPATGKPE 2317

Qy     151 AQMMWYKNSSLLDLEKSRHQIQQTS-ESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 209
             : |:|:   : :|     |        || ||:::: | | |:| | :|  :   | :
Db    2318 PDITWFKDGEAIHIENIADIIPNGELNGNQLKIIRIKEGDAGKYICEADNSAGSVEQDVN 2377

Qy     210 L----IVKDKTVGVPI-------EPIIAACVV 230
           :    | | :  |:|        | :: :| |
Db    2378 VNVITIPKIEKDGIPSDYESQQNERVVISCPV 2409
```

RESULT 15
T43290
hemicentin precursor - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C;Accession: T43290; T20993; T24734
R;Vogel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A;Description: Hemicentin is required for hemidesmosome mediated cell adhesion and ge
A;Reference number: Z22396
A;Accession: T43290
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5198 <VOG>
A;Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R;Sulston, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19355
A;Accession: T20993
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5198 <WIL>
A;Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b
A;Experimental source: clone F15G9
R;Kershaw, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19929
A;Accession: T24734
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5198 <WI2>
A;Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b
A;Experimental source: clone T09B9
C;Genetics:
A;Gene: him-4; F15G9.4b

A;Map position: X
A;Introns: 8571; 13871; 44471; 47371; 47771; 52671; 56471; 57771; 59171; 11847/3;
; 25127; 27647; 28097; 28771; 28971; 29037; 29047; 29417; 29637; 31337/1
1; 42257/1; 43617/1; 44087/1; 44567/1; 44987/1; 46477/3; 48387/1; 49027/1; 49647/1; 50347/1; 5109

Query Match 11.4%; Score 152.5; DB 2; Length 5198;
Best Local Similarity 23.1%; Pred. No. 0.0058;
Matches 49; Conservative 42; Mismatches 98; Indels 23; Gaps 8;

```
Qy      41 ILDTTPGSQASLICAVQNHTK-EEELLWYREEGRVDLKSGNK-----INSSSVCVSSISE 94
           |:    |       | : :    : :::| |    :||:: :       |   : : :::|
Db    2199 IVTAIKGGALPFKCPIDDDKNFKGQIIWLRNYQPIDLEAEDARITRLSNDRRLTILNVTE 2258

Qy      95 NDNGISFTCRLGRDQSV-SVSVVLNVTFPP---LLSGNDFQTVEEGSNVKLVCNVKANPQ 150
           || |  ::||:  |    |       |  ||    :|  :  :|  | | | | |     |:
Db    2259 NDEG-QYSCRVKNDAGENSFDFKATVLVPPTIIMIDKDKNKTAVFHSTVTLSCPATGKPE 2317

Qy     151 AQMMWIKNSSLLDLEKSRHQTQQTS-ESFQLSTIKVEKPDNGTYSCIAKSSLKTESLDFH 209
             : |:|:   : :|     |         || ||:::: | | |:| | :|  :   | :
Db    2318 PDITWFKDGEAIHLENIADTIPNGELNGNQLKTIKRAGDAGKYTCEADNSAGSVEQDVN 2377

Qy     210 L----IVKDKTVGVPI-------EPIIAACVV 230
           :    | | :  |:|         | :: :| |
Db    2378 VNVITIPKIEKDGIPSDYESQQNERVVISCPV 2409
```

Search completed: September 17, 2002, 13:12:01
Job time: 11233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:10:53 ; Search time 40.16 Seconds
(without alignments)
244.890 Million cell updates/sec

Title: US-09-863-823-2
Perfect score: 1295
Sequence: 1 MQMGRFLLLVILFLPREMTS.....RKKIMKLCMKDKDPHSETAL 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	176	13.6	853	1	NCA1_BOVIN	P31836 bos taurus
2	175	13.5	858	1	NCA1_RAT	P13596 rattus norv
3	170	13.1	2012	1	DSCA_HUMAN	O60469 homo sapien
4	169.5	13.1	359	1	LACH_DROME	Q24372 drosophila
5	166	12.8	761	1	NCA2_HUMAN	P13592 homo sapien
6	166	12.8	848	1	NCA1_HUMAN	P13591 homo sapien
7	160.5	12.4	1091	1	NCA1_CHICK	P13590 gallus gall
8	159	12.3	725	1	NCA2_MOUSE	P13594 mus musculu
9	159	12.3	1115	1	NCA1_MOUSE	P13595 mus musculu
10	153	11.8	2029	1	LAR_DROME	P16621 drosophila
11	149	11.5	1447	1	DCC_HUMAN	P43146 homo sapien
12	149	11.5	1912	1	PTPD_HUMAN	P23468 homo sapien
13	148.5	11.5	338	1	LAMP_HUMAN	Q13449 homo sapien
14	147.5	11.4	4393	1	PGBM_HUMAN	P98160 homo sapien
15	147	11.4	1447	1	DCC_MOUSE	P70211 mus musculu
16	145	11.2	837	1	NCM2_HUMAN	O15394 homo sapien
17	144.5	11.2	338	1	LAMP_CHICK	Q98919 gallus gall
18	144.5	11.2	338	1	LAMP_RAT	Q62813 rattus norv
19	144.5	11.2	1348	1	VGR2_COTJA	P52583 coturnix co
20	144	11.1	521	1	CEA1_MOUSE	P31809 mus musculu
21	142.5	11.0	349	1	LACH_SCHAM	Q26474 schistocerc
22	142	11.0	837	1	NCM2_MOUSE	O35136 mus musculu
23	142	11.0	1356	1	VGR2_HUMAN	P35968 homo sapien
24	141.5	10.9	1493	1	NEO1_MOUSE	P97798 mus musculu
25	139	10.7	1450	1	MPSF_CHICK	Q02173 gallus gall
26	137	10.6	1092	1	NCA2_XENLA	P36335 xenopus lae
27	136.5	10.5	519	1	ECTO_RAT	P16573 rattus norv
28	136.5	10.5	1343	1	VGR2_RAT	O08775 rattus norv
29	136.5	10.5	1897	1	PTPF_HUMAN	P10586 homo sapien
30	135.5	10.5	1070	1	PTK7_HUMAN	Q13308 homo sapien
31	133.5	10.3	344	1	CEA6_HUMAN	P40199 homo sapien
32	133.5	10.3	3707	1	PGBM_MOUSE	Q05793 mus musculu
33	132.5	10.2	976	1	KIT_HUMAN	P10721 homo sapien
34	132	10.2	1377	1	NEO1_RAT	P97603 rattus norv
35	131.5	10.2	353	1	CEPU_CHICK	Q90773 gallus gall
36	131.5	10.2	526	1	CEA1_HUMAN	P13688 homo sapien
37	131	10.1	1088	1	NCA1_XENLA	P16170 xenopus lae
38	130.5	10.1	298	1	JAM2_HUMAN	P57087 homo sapien
39	130.5	10.1	702	1	CEA5_HUMAN	P06731 homo sapien
40	130.5	10.1	864	1	TRKC_RAT	Q03351 rattus norv
41	130	10.0	1367	1	VGR2_MOUSE	P35918 mus musculu
42	129.5	10.0	626	1	MAG_MOUSE	P20917 mus musculu
43	129.5	10.0	822	1	FGR1_HUMAN	P11362 homo sapien
44	129.5	10.0	822	1	FGR1_MOUSE	P16092 mus musculu
45	129.5	10.0	822	1	FGR1_RAT	Q04589 rattus norv

ALIGNMENTS

RESULT 1
NCA1_BOVIN
ID NCA1_BOVIN STANDARD; PRT; 853 AA.
AC P31836;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE (NCAM-140).
GN NCAM1 OR NCAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain cortex;
RX MEDLINE=89378239; PubMed=2776887;
RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshnyakov M.V.,
RA Petukhova G.V., Rakitina T.V., Feshchenko E.A., Ishchenko K.A.,
RA Mirzoeva S.F., Chernova M.N., Dranytsyna S.M.;
RT "Calmodulin-independent bovine brain adenylate cyclase. Amino acid
RT sequence and nucleotide sequence of the corresponding cDNA.";
RL FEBS Lett. 254:69-73(1989).
RN [2]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401(1986).
RN [3]
RP IDENTIFICATION AS N-CAM.
RX MEDLINE=92111748; PubMed=1765159;
RA Premont R.T.;
RT "A bovine brain cDNA purported to encode calmodulin-insensitive
RT adenylyl cyclase has extensive identity with neural cell adhesion
RT molecules (N-CAMs).";
RL FEBS Lett. 295:230-231(1991).
CC -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- CAUTION: WAS ORIGINALLY (REF 1) THOUGHT TO BE A CALMODULIN-
CC INDEPENDENT ADENYLATE CYCLASE.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; X16451; CAA34470.1; -.
DR   PIR; A32976; IJBONC.
DR   HSSP; P40189; 1BQU.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20    858       NEURAL CELL ADHESION MOLECULE, 140 KDA
FT                                 ISOFORM.
FT   DOMAIN       20    719       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    720    737       POTENTIAL.
FT   DOMAIN      738    858       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    293       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      321    401       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      428    495       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      527    604       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      633    700       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       BY SIMILARITY.
FT   DISULFID    139    189       BY SIMILARITY.
FT   DISULFID    235    286       BY SIMILARITY.
FT   DISULFID    328    394       BY SIMILARITY.
FT   DISULFID    435    488       BY SIMILARITY.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    314    314       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    346    346       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    458    458       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   853 AA;  93893 MW;  E12FD49231A7A368 CRC64;


  Query Match             13.6%;  Score 176;  DB 1;  Length 853;
  Best Local Similarity   28.4%;  Pred. No. 6.7e-07;
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      41 QASLIC--------------AVQNHTREEELLWYRKEDGRVDLESGNKTNS----SSVLVS 82
           | :| |                : :| : ||:  | | | :  |     : |    ||: :
Db     323 QVTLTCEASGDPIPSITWRTSTRNISSEEKASWTRPEKQETLDGHMVVRSHARVSSLTLK 382

Qy      83 SISENDNGTSFIC    RLGRDQDSVSVSVVLNVTFPPILSGNDFQTVEEGSNVKLVCNVK 138
           ||   | | : :|     :|:|   | |: | | : | | |     ||: | : | |
Db     383 SIQYTDAG-EYVCTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 438

Qy     139 ANPQAQMMWYKNSSLLDLEK-SRRQTQQTSFSFQLSITKVEKPDNGTYSCIAKSSLATES 197
           | | | : |::   ||     |  :|  |  :  | :|   : | | |:| | : :  ||
Db     439 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 498

Qy     198 LDFHLIVFDKTVGVP----IFPLIAACVVIF 224
           |:| |:  | |   |    :|| :   | |
Db     499 LEFVLVQAD-TPSSPSIDQVEPYSSTAQVQF 528


RESULT   2
NCA1_RAT
ID   NCA1_RAT       STANDARD;      PRT;   858 AA.
AC   P13596;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE   (NCAM-140).
GN   NCAM1 OR NCAM.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=88059265; PubMed=3680385;
RA   Small S.J., Shull G.E., Santoni M.-J., Akeson R.;
RT   "Identification of a cDNA clone that contains the complete coding
RT   sequence for a 140 kD rat NCAM polypeptide.",
RL   J. Cell Biol. 105:2335-2345(1987).
RN   [2]
RP   SEQUENCE OF 355-364 FROM N.A.
RX   MEDLINE=90166485; PubMed=2483093;
RA   Small S.J., Haines S.L., Akeson R.A.;
RT   "Polypeptide variation in an N-CAM extracellular immunoglobulin-like
RT   fold is developmentally regulated through alternative splicing.";
RL   Neuron 1:1007-1017(1988).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; X06564; CAA29809.1; -.
DR   EMBL; M32611; AAA41679.1; -.
DR   PIR; S00846; IJRTNC.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.
FT   SIGNAL        1     19
FT   CHAIN        20    858       NEURAL CELL ADHESION MOLECULE, 140 KDA
FT                                 ISOFORM.
FT   DOMAIN       20    721       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    722    739       POTENTIAL.
FT   DOMAIN      740    858       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    295       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      323    403       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      430    497       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      514    615       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      616    712       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       BY SIMILARITY.
FT   DISULFID    139    189       BY SIMILARITY.
FT   DISULFID    235    288       BY SIMILARITY.
FT   DISULFID    330    396       BY SIMILARITY.
FT   DISULFID    437    490       BY SIMILARITY.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    348    348       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    434    434       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    460    460       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   858 AA;  94658 MW;  EA1A06A4EA0550F6 CRC64;


  Query Match             13.5%;  Score 175,  DB 1,  Length 858,
  Best Local Similarity   28.4%;  Pred. No. 8.2e-07;
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      41 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINS----SSVCVS 82
           | :| |              : :| : ||:  | | | :  |     ' !    ||' '
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEKASWTRPEKQETLDGHMVVRSHARVSSLTLK 384

Qy      83 SISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 138
           ||   | |  : |     :|:|   | |: } | : | | |        ||: | : | |
Db     385 SIQYTDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 440

Qy     139 ANPQAQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTES 197
           | | | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  ||
Db     441 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 500

Qy     198 LDFHLIVKDKTVGVP----IEPIIAACVVIF 224
           |:| |:  | |   |    :|| :   | |
Db     501 LEFILVQAD-TPSSPSIDRVEPYSSTAQVQF 530


RESULT   3
DSCA_HUMAN
ID   DSCA_HUMAN     STANDARD;      PRT;  2012 AA.
AC   O60469; O60468;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Down syndrome cell adhesion molecule precursor (CHD2).
GN   DSCAM.
OS   Homo sapiens (Human).
OC   Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
OC   Mammalia, Eutheria, Primates, Catarrhini, Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=98087574; PubMed=9426258;
RA   Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA   Lyons G.E., Korenberg J.R.;
RT   "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT   Down syndrome region and is involved in the development of the
RT   nervous system.";
RL   Hum. Mol. Genet. 7:227-237(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=20384934; PubMed=10925149;
RA   Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT   "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT   intercellular adhesion.";
RL   Brain Res. Mol. Brain Res. 79:118-126(2000).
CC   -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC       INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC       NERVOUS SYSTEM DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC       SHORT ISOFORM MAY BE SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/CHD2-52 (SHOWN HERE)
CC       AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF023450; AAC17967.1; -.
DR   EMBL; AF023449; AAC17966.1; -.
DR   EMBL; AL163283; CAB90464.1; -.
DR   EMBL; AL163282; CAB90436.1; -.
DR   EMBL; AL163281; CAB90444.1; -.
DR   HSSP; P40189; 1BQU.
DR   MIM; 602523; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 9.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 7.
KW   Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW   Transmembrane; Alternative splicing.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   2012       DOWN SYNDROME CELL ADHESION MOLECULE.
FT   DOMAIN       18   1595       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1596   1616       POTENTIAL.
FT   DOMAIN     1617   2012       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39    109       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      138    204       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      239    300       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      328    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      421    491       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    582       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      610    676       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      704    773       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      802    872       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN      885    972       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      984   1076       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     1088   1177       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     1189   1273       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     1300   1366       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     1380   1463       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     1477   1562       FIBRONECTIN TYPE-III 6.
FT   DISULFID     46    102       BY SIMILARITY.
FT   DISULFID    145    197       BY SIMILARITY.
FT   DISULFID    246    293       BY SIMILARITY.
FT   DISULFID    335    385       BY SIMILARITY.
FT   DISULFID    428    484       BY SIMILARITY.
FT   DISULFID    525    575       BY SIMILARITY.
FT   DISULFID    617    669       BY SIMILARITY.
FT   DISULFID    711    766       BY SIMILARITY.
FT   DISULFID    809    865       BY SIMILARITY.
```

```
FT   DISULFID   1307   1359       BY SIMILARITY.
FT   CARBOHYD     28     28       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     78     78       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    512    512       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    748    748       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    795    795       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1142   1142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1160   1160       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1250   1250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1271   1271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1341   1341       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1488   1488       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1562   1571       NFATLNYIGS -> KEAARCKEFS (IN SHORT
FT                                ISOFORM).
FT   VARSPLIC   1572   2012       MISSING (IN SHORT ISOFORM).
FT   CONFLICT   1893   2012       HRPGDLIHLPPYLRMDPLLNRGGPGTSRDLSLGQACLEPQK
FT                                SRTLKRPTVLEPIPMEAASSASSTREGQSWQPGAVATLPQR
FT                                EGAELGQAAKMSSSQESLLDSRGHLKGNNPYAKSYTLV ->
FT                                IGQVTSYICLHTLEWTFC (IN REF. 1).
SQ   SEQUENCE   2012 AA;  222259 MW;  0E33CFB781A08334 CRC64;


  Query Match             13.1%;  Score 170;  DB 1;  Length 2012;
  Best Local Similarity   29.8%;  Pred. No. 6.3e-06;
  Matches   57;  Conservative   31;  Mismatches   69;  Indels   34;  Gaps   10;

Qy      15 PREMTSSVILTVNGKTFNYILDTTPGSQASLICAVQNHTREEPFTLWYREEGRVIDLKSGNK- 73
           ||:: |||                 ||| || |:|   | ::|| |||  | : |  |
Db     320 PRKVKSSV----------------GSQVSLSCSVIG-TEDQELSWYR-NGEI-LNPGKNV 360

Qy      74 - - INSSSVCVSSISENDNGISFTCRLGRDQSVS   VSVVLNVTFPPTLSGNDFQTVE 126
             ||  :: :  : ::| | :: | : :|:  :    | |||     | ::|    : |
Db     361 RITGINHENLIMDHMVKSDGG-AYQCEVRKDKLSAQDYVQVVLEDGTPKIISAFSEKVVS 419

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSERQIQQ--ISES---FQLSITKVEKPD 181
           | |:|||  |   : |  :   : |:   |:| |  |||      |:|:  :  |
Db     420 PAEPVSLMCNVKGTPLPTITWTLDDDPI-LKGGSHRISQMITSEGNVVSYLNISSSQVRD 478

Qy     182 NGTYSCIAKSS 192
           | | | | :|
Db     479 GGVYRCIANNS 489


RESULT   4
LACH_DROME
ID   LACH_DROME     STANDARD;      PRT;   359 AA.
AC   Q24372;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Lachesin precursor.
GN   LAC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pteryqota; Neoptera; Endopteryqota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94038693; PubMed=8223276;
RA   Karlstrom R.O., Wilder I.P., Bastiani M.J.;
RT   "Lachesin: an immunoglobulin superfamily protein whose expression
RT   correlates with neurogenesis in grasshopper embryos.";
RL   Development 118:509-522(1993).
CC   -!- FUNCTION: MAY PLAY A ROLE IN EARLY NEURONAL DIFFERENTIATION AND
CC       AXON OUTGROWTH.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ON DIFFERENTIATING NEURONAL CELLS
CC       FROM THE ONSET OF NEUROGENESIS IN BOTH THE CENTRAL AND PERIPHERAL
CC       NERVOUS SYSTEMS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13255; AAC37184.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0010238; Lac.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    336       LACHESIN.
FT   PROPEP      337    359       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       43    120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      150    211       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      240    310       IG-LIKE C2-TYPE DOMAIN 2.
FT   DISULFID     50    113       POTENTIAL.
FT   DISULFID    157    204       POTENTIAL.
FT   DISULFID    247    303       POTENTIAL.
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       336    336       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   359 AA;  39998 MW;  D1F513E2B9D886E6 CRC64;


  Query Match             13.1%;  Score 169.5;  DB 1;  Length 359;
  Best Local Similarity   27.1%;  Pred. No. 7.6e-07;
  Matches   56;  Conservative   42;  Mismatches   78;  Indels   31;  Gaps   10;

Qy       8 LLVILELPREMISSVLIVNGKIENYILDTTPGSQASLICAVQNHTREEELLWYR-EEGRV 66
           ||: :|: : :     |:: |:  | |   |      |:|| : :|  :|: : :   |
Db      14 LLLAIFVQQTLAQRTPTISYITQEQIKDI--GGTVEFDCSVQ-YAKEYNVLFLKTDSDPV 70

Qy      67 DLKSGNK!-------------NSSS--VCVSSISENDNGISFTCR--LGRDQSVSVSVVL 109
             | :|: :              |||:  : :  | | | | ::|||:  :     ||  | |
Db      71 FLSIGSILVIKDSREDLRYDPNSSFYKDGIEIDQEIDAG-IYICQVVISIVHRVSAEVKL 129

Qy     110 NVTFPPTLSGNDFQTV- EEGSNVKLVCNVKANPQAQMMWYR-NSSLLDLEKSRQIQQT 166
           :|  ||::| |  |:|   ||| |:: |     |  : : | |:::|  : :    |
Db     130 SVRRPPVISDNSTQSVVASEGSEVQMECYASGYPTPTITWRRENNAILPTDSA------T 183

Qy     167 SESFQLSITKVEKPDNGTYSCIAKSSL 193
             | |  |:| | ||| |:| : :
Db     184 YVGNTLRIKSVKKEDRGTYYCVADNGV 210


RESULT   5
NCA2_HUMAN
ID   NCA2_HUMAN     STANDARD;      PRT;   761 AA.
AC   P13592; P13593;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, phosphatidylinositol-linked isoform
DE   precursor (N-CAM 120) (NCAM-120) (CD56 antigen).
```

```
GN   NCAM1 OR NCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (N-CAM 120).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=89305258; PubMed=3253057;
RA   Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,
RA   Elsom V., Moore S.E., Goridis C., Walsh F.S.;
RT   "Complete sequence and in vitro expression of a tissue-specific
RT   phosphatidylinositol-linked N-CAM isoform from skeletal muscle.";
RL   Development 104:165-173(1988).
RN   [2]
RP   SEQUENCE OF 491-761 FROM N.A. (N-CAM 120).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=87301755; PubMed=2887295;
RA   Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA   Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT   "Human muscle neural cell adhesion molecule (N-CAM): identification
RT   of a muscle-specific sequence in the extracellular domain.";
RL   Cell 50:1119-1130(1987).
RN   [3]
RP   SEQUENCE OF 491-655 FROM N.A. (SECRETED ISOFORM).
RX   MEDLINE=89077552; PubMed=3203385;
RA   Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,
RA   Dickson G., Walsh F.S.;
RT   "Alternative splicing generates a secreted form of N-CAM in muscle
RT   and brain.";
RL   Cell 55:955-964(1988).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16841; CAA34739.1; -.
DR   EMBL; M17409; AAA59912.1; -.
DR   EMBL; M22094; AAA59910.1; -.
DR   EMBL; M22092; AAA59911.1; -.
DR   EMBL; M22091; AAA59911.1; JOINED.
DR   PIR; S07784; IJHUNG.
DR   PIR; A31635; A31635.
DR   MIM; 116930; -.
DR   InterPro; IPR003961, FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW   GPI-anchor; Alternative splicing.
FT   SIGNAL       1     19
FT   CHAIN       20    761       NEURAL CELL ADHESION MOLECULE,
FT                                PHOSPHATIDYLINOSITOL-LINKED ISOFORM.
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    294       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      322    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      419    486       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    595       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      660    727       FIBRONECTIN TYPE-III 2.
FT   DISULFID     41     96       PROBABLE.
FT   DISULFID    139    189       PROBABLE.
FT   DISULFID    235    287       PROBABLE.
FT   DISULFID    329    385       PROBABLE.
FT   DISULFID    426    479       PROBABLE.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    635    655       GEPSAPKLEGQMGEDGNSIKV -> NIAQNHCCNMFQAGLH
FT                                NALMK (IN SECRETED ISOFORM).
SQ   SEQUENCE   761 AA;  83770 MW;  F0CAD3292D7AB67E CRC64;


  Query Match             12.8%;  Score 166;  DB 1;  Length 761;
  Best Local Similarity   28.0%;  Pred. No. 3.9e-06;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      41 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 86
           | :| |                : :| : ||: |    :| : ::|  ::  ||: : ||
Db     324 QVTLTCEASGDPIPSITWRTSTRNISSEEKTL----DGHMVVRSHARV--SSLTLKSIQY 377

Qy      87 NDNGISFTC----KLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 142
             | |  : |     :|:|    | |: | | : | | |        ||: | : | | | |
Db     378 TDAG-EYTCTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 433

Qy     143 AQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTFSLDFH 201
           | : |::  ||     |  :| |  :  | :|   : | | |:| | : :  |||:|
Db     434 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 493

Qy     202 LIVKDKTVGVP----IEPIIAACVVIF 224
           |:  | |   |    :|| :  | |
Db     494 LVQAD-TPSSPSIDQVEPYSSTAQVQF 519


RESULT   6
NCA1_HUMAN
ID   NCA1_HUMAN     STANDARD;      PRT;   848 AA.
AC   P13591; Q16180; Q15829;
DT   01-JAN-1990 (Rel. 13, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE   (NCAM-140) (CD56 antigen).
GN   NCAM1 OR NCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94356433; PubMed=8075973;
RA   Saito S., Ianio Y., Iachibana I., Hayashi S., Kishimoto T., Kawase I.;
RT   "Complementary DNA sequence encoding the major neural cell adhesion
RT   molecule isoform in a human small cell lung cancer cell line.";
RL   Lung Cancer 10:307-318(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91250739; PubMed=1710251;
RA   Lanier L.L., Chang C., Azuma M., Ruitenberg J.J., Hemperly J.J.,
RA   Phillips J.H.;
RT   "Molecular and functional analysis of human natural killer cell-
RT   associated neural cell adhesion molecule (N-CAM/CD56).";
RL   J. Immunol. 146:4421-4426(1991).
RN   [3]
RP   SEQUENCE OF 491-848 FROM N.A.
```

```
RX   MEDLINE=87301755; PubMed=2887295;
RA   Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA   Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT   "Human muscle neural cell adhesion molecule (N-CAM): identification
RT   of a muscle specific sequence in the extracellular domain.";
RL   Cell 50:1119-1130(1987).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC    !- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS
CC    !- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD quide CD5b entry;
CC       WWW-"http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S71824; AAB31836.1; -.
DR   EMBL; U64041; AAB04558.1; -.
DR   EMBL; M17410; AAA59913.1; -.
DR   PIR; B26883; B26883.
DR   HSSP; P40189; 1BQU.
DR   MIM; 116930; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW   Transmembrane, Alternative splicing.
FT   SIGNAL       1     19
FT   CHAIN       20    848       NEURAL CELL ADHESION MOLECULE, 140 KDA
FT                               ISOFORM.
FT   DOMAIN       20    708       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    709    729       POTENTIAL.
FT   DOMAIN      730    848       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    294       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      322    392       IG LIKE C2 TYPE DOMAIN 4.
FT   DOMAIN      419    486       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    595       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      660    727       FIBRONECTIN TYPE-III 2.
FT   DISULFID     41     96       PROBABLE.
FT   DISULFID    139    189       PROBABLE.
FT   DISULFID    235    287       PROBABLE.
FT   DISULFID    329    385       PROBABLE.
FT   DISULFID    426    479       PROBABLE.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL)
FT   CARBOHYD    347    347       N-LINKED (GLCNAC. . .) (POTENTIAL)
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    215    215       Q -> R (IN REF. 2).
FT   CONFLICT    239    239       G -> R (IN REF. 2).
FT   CONFLICT    490    490       L -> F (IN REF. 2).
FT   CONFLICT    599    600       QG -> R (IN REF. 3).
FT   CONFLICT    720    721       MISSING (IN REF. 3).
FT   CONFLICT    811    811       G -> A (IN REF. 3).
SQ   SEQUENCE   848 AA;  93360 MW;  68D2F000E6C1C2AD CRC64;
```

```
  Query Match             12.8%;  Score 166;  DB 1;  Length 848;
  Best Local Similarity   28.0%;  Pred. No. 4.4e-06;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      41 QASIC--------------AVQNRTPEFEEIIWYPEEGRVIQLKSGNKINSSSVCVSSISE 86
           | :| |               : :| : ||: |    :| : ::|  ::  |||: : ||
Db     324 QVILICEASGDPIPSITWRISTRNISSEEKTL----DGHMVVRSHARV--SSLTLKSIQY 377

Qy      87 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 142
           | |  : |      :|:|   | |: | | : | | |          ||: | : | | | |
Db     378 IDAG-EYICIASNTIGQD---SQSMYLEVQYAPELQGPVAVYTWEGNQVNITCEVFAYPS 433

Qy     143 AQMMWYFNES----DEF-SRDQLPDSEFQGRFRVFKLDNGIYSCIARSSTFTFSLDFH 201
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     434 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 493

Qy     202 LIVKDKIVGVP----IEPIIAACVVIF 224
           |:  | |   |    :|| :   | |
Db     494 LVQAD-TPSSPSIDQVEPYSSTAQVQF 519


RESULT   7
NCAL_CHICK
ID   NCAL_CHICK     STANDARD;      PRT;  1091 AA.
AC   P13590; Q90919; Q90918;
DT   01-JAN-1990 (Rel. 13, Created)
DT   15 JUL 1998 (Rel. 36, Last sequence update)
DT   01 MAR 2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, large isoform precursor (N-CAM 180)
DE   (Contains: N-CAM 140).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-175 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=87206190; PubMed=3576199;
RA   Cunningham B.A., Hemperly J.J., Murray B.A., Prediger E.A.,
RA   Brackenbury R., Edelman G.M.;
RT   "Neural cell adhesion molecule: structure, immunoglobulin-like
RT   domains, cell surface modulation, and alternative RNA splicing.";
RL   Science 236:799-806(1987).
RN   [2]
RP   SEQUENCE OF 128-1091 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=86206089; PubMed=3458261;
RA   Hemperly J.J., Murray B.A., Edelman G.M., Cunningham B.A.;
RT   "Sequence of a cDNA clone encoding the polysialic acid-rich and
RT   cytoplasmic domains of the neural cell adhesion molecule N-CAM.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:3037-3041(1986).
RN   [3]
RP   SEQUENCE FROM N.A. (726 AA FORM).
RX   MEDLINE=87092340; PubMed=3467341;
RA   Hemperly J.J., Edelman G.M., Cunningham B.A.;
RT   "cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a
RT   membrane spanning region consistent with evidence for membrane
RT   attachment via a phosphatidylinositol intermediate.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:9822-9826(1986).
RN   [4]
RP   SEQUENCE OF 810-1069 FROM N.A.
RX   MEDLINE=87033934; PubMed=3771645;
RA   Murray B.A., Owens G.C., Prediger E.A., Crossin K.L.,
RA   Cunningham B.A., Edelman G.M.;
RT   "Cell surface modulation of the neural cell adhesion molecule
RT   resulting from alternative mRNA splicing in a tissue-specific
RT   developmental sequence.";
RL   J. Cell Biol. 103:1431-1439(1986).
RN   [5]
RP   SEQUENCE OF 1-17 FROM N.A.
RX   MEDLINE=93122797; PubMed=1478668;
RA   Colwell G., Li B., Forrest D., Brackenbury R.;
RT   "Conserved regulatory elements in the promoter region of the N-CAM
```

```
RT   gene.";
RL   Genomics 14:875-882(1992).
RN   [6]
RP   SEQUENCE OF 1-17 FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=Erythrocyte;
RA   Sasner M., Covault J.;
RL   Submitted (AUG-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15861; AAB59958.1; -.
DR   EMBL; M15860; AAB59958.1; JOINED.
DR   EMBL; M15922; AAB59958.1; JOINED.
DR   EMBL; M15923; AAB59958.1; JOINED.
DR   EMBL; M15924; AAB59958.1; JOINED.
DR   EMBL; M21178; AAB59958.1; JOINED.
DR   EMBL; M21179; AAB59958.1; JOINED.
DR   EMBL; M21180; AAB59958.1; JOINED.
DR   EMBL; M15929; AAB59958.1; JOINED.
DR   EMBL; M15930; AAB59958.1; JOINED.
DR   EMBL; M15931; AAB59958.1; JOINED.
DR   EMBL; M15932; AAB59958.1; JOINED.
DR   EMBL; M15933; AAB59958.1; JOINED.
DR   EMBL; M15934; AAB59958.1; JOINED.
DR   EMBL; L29437; AAB59958.1; JOINED.
DR   EMBL; M15935; AAB59958.1; JOINED.
DR   EMBL; M15937; AAB59958.1; JOINED.
DR   EMBL; M15938; AAB59958.1; JOINED.
DR   EMBL; M15939; AAB59958.1; JOINED.
DR   EMBL; M15861; AAB59959.1; -.
DR   EMBL; M15860; AAB59959.1; JOINED
DR   EMBL; M15922; AAB59959.1; JOINED.
DR   EMBL; M15923; AAB59959.1; JOINED
DR   EMBL; M15924; AAB59959.1; JOINED.
DR   EMBL; M21178; AAB59959.1; JOINED.
DR   EMBL; M21179; AAB59959.1; JOINED.
DR   EMBL; M21180; AAB59959.1; JOINED.
DR   EMBL; M15929; AAB59959.1; JOINED.
DR   EMBL; M15930; AAB59959.1; JOINED.
DR   EMBL; M15931; AAB59959.1; JOINED.
DR   EMBL; M15932; AAB59959.1; JOINED.
DR   EMBL; M15934; AAB59959.1; JOINED.
DR   EMBL; L29437; AAB59959.1; JOINED.
DR   EMBL; M15935; AAB59959.1; JOINED.
DR   EMBL; M15937; AAB59959.1; JOINED.
DR   EMBL; M15939; AAB59959.1; JOINED.
DR   EMBL; M15936; AAB59957.1; -.
DR   EMBL; M15860; AAB59957.1; JOINED.
DR   EMBL; M15922; AAB59957.1; JOINED.
DR   EMBL; M15923; AAB59957.1; JOINED.
DR   EMBL; M15924; AAB59957.1; JOINED.
DR   EMBL; M21178; AAB59957.1; JOINED.
DR   EMBL; M21179; AAB59957.1; JOINED.
DR   EMBL; M21180; AAB59957.1; JOINED.
DR   EMBL; M15929; AAB59957.1; JOINED
DR   EMBL; M15930; AAB59957.1; JOINED.
DR   EMBL; M15931; AAB59957.1; JOINED.
DR   EMBL; M15932; AAB59957.1; JOINED.
DR   EMBL; M15934; AAB59957.1; JOINED.
DR   EMBL; L29437; AAB59957.1; JOINED.
DR   EMBL; M15935; AAB59957.1; JOINED.
DR   EMBL; X04479; CAB51638.1; -.
DR   EMBL; X70342; CAA49807.1; -.
DR   EMBL; Z12128; CAA78113.1; -.
DR   PIR; A43613; IJCHNL.
DR   HSSP; Q13740; 1KJC.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 4.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20   1091       NEURAL CELL ADHESION MOLECULE, LARGE
FT                                ISOFORM.
FT   DOMAIN       20    711       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    712    729       POTENTIAL.
FT   DOMAIN      730   1091       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    294       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      322    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      419    486       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    595       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      624    692       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       BY SIMILARITY.
FT   DISULFID    139    189       BY SIMILARITY.
FT   DISULFID    235    287       BY SIMILARITY.
FT   DISULFID    329    385       BY SIMILARITY.
FT   DISULFID    426    479       BY SIMILARITY.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    702    726       STSPTSGLGTAAIVGILIVIFVLIL -> TLGSPSTSSSFV
FT                                SLLLSAVTLLLLC (IN 726 AA ISOFORM).
FT   VARSPLIC    727   1091       MISSING (IN 726 AA ISOFORM).
FT   VARSPLIC    771    809       MISSING (IN 1052 AA ISOFORM).
FT   VARSPLIC    810   1070       MISSING (IN ISOFORM N-CAM 140).
SQ   SEQUENCE   1091 AA;  117415 MW;  B151A67002DF88E0 CRC64;


  Query Match             12.4%;  Score 160.5;  DB 1;  Length 1091;
  Best Local Similarity   26.9%;  Pred. No. 1.7e-05;
  Matches   60;  Conservative   37;  Mismatches   87;  Indels   39;  Gaps   10;

Qy      25 VNGKTENYILDTTPGSQASLIC-------------AVQNHTREEELLWYREEGRVDLKS 70
           |  ||    : |      | :| |                 : :| : ||: |    :||: ::|
Db     313 VENKTAMELED-----QITLTCEASGDPIPSITWKTSTRNISNEEKTL----DGRIVVRS 363

Qy      71 GNKINSSSVCVSSISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVE 126
             ::  ||: :  |   | |  : |      :|:|   | :: | | : | | |
Db     364 HARV--SSLTLKEIQYTDAG-EYVCTASNTIGQD---SQAMYLEVQYAPKLQGPVAVYTW 417

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTY 185
           ||: | : | | | | | : |::  ||     |  :|  |  :  | :|   : | | |
Db     418 EGNQVNITCEVFAYPSAVISWFRDGQLLPSSNYSNIKIYNTPSASYLEVIPDSENDFGNY 477

Qy     186 SCIAKSSLKTESLDFHLIVKDKTVGVP----IEPIIAACVVIF 224
           :| | : :  || :| |:  | |   |     :|| :   | |
Db     478 NCTAVNRIGQESSEFILVQAD-TPSSPSIDRVEPYSSTARVEF 519
```

```
RESULT   8
NCA2_MOUSE
ID   NCA2_MOUSE      STANDARD;      PRT:   725 AA.
AC   P13594; Q61950;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, phosphatidylinositol-linked isoform
DE   precursor (N-CAM 120) (NCAM-120).
GN   NCAM1 OR NCAM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=87246524; PubMed=3595563;
RA   Barthels D., Santoni M.-J., Wille W., Ruppert C., Caix J.-C.,
RA   Hirsch M.-R., Fontecilla-Camps J.-C., Goridis C.;
RT   "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT   a Mr 79,000 polypeptide without a membrane-spanning region.";
RL   EMBO J. 6:907-914(1987).
RN   [2]
RP   SEQUENCE OF 20-700 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=89251563; PubMed=2721486;
RA   Santoni M.J., Barthels D., Vopper G., Boned A., Goridis C., Wille M.;
RT   "Differential exon usage involving an unusual splicing mechanism
RT   generates at least eight types of NCAM cDNA in mouse brain.";
RL   EMBO J. 8:385-392(1989).
RN   [3]
RP   SEQUENCE OF 642-725 FROM N.A.
RX   MEDLINE=88284628; PubMed=3396534;
RA   Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.;
RT   "Differential splicing and alternative polyadenylation generates
RT   distinct NCAM transcripts and proteins in the mouse.";
RL   EMBO J. 7:625-632(1988).
RN   [4]
RP   SEQUENCE OF 20-36.
RX   MEDLINE=86140120; PubMed=3512556;
RA   Rougon G., Marshak D.R.;
RT   "Structural and immunological characterization of the amino-terminal
RT   domain of mammalian neural cell adhesion molecules.";
RL   J. Biol. Chem. 261:3396-3401(1986).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y00051; CAA68263.1; -.
DR   EMBL; X15049; CAA33148.1; ALT_SEQ.
DR   EMBL; X07195; CAA30173.1; -.
DR   PIR; A29673; IJMSNC.
DR   MGD; MGI:97281; Ncam.
DR   InterPro; IPR003961; FN_III
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
KW   Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor.
FT   SIGNAL        1     19
FT   CHAIN        20    725       NEURAL CELL ADHESION MOLECULE,
FT                                 PHOSPHATIDYLINOSITOL-LINKED ISOFORM.
FT   DOMAIN        34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN       132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN       228    295       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN       323    393       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN       420    487       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN       519    596       FIBRONECTIN TYPE-III 1.
FT   DOMAIN       625    692       FIBRONECTIN TYPE-III 2.
FT   DOMAIN       152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN       161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID      41     96       PROBABLE.
FT   DISULFID     139    189       PROBABLE.
FT   DISULFID     235    288       PROBABLE.
FT   DISULFID     330    386       PROBABLE.
FT   DISULFID     427    480       PROBABLE.
FT   CARBOHYD     222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     348    348       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     424    424       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     450    450       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     479    479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     261    268       ERSRSSVS -> DEKHIFSD (IN REF. 2).
FT   CONFLICT     273    273       V -> L (IN REF. 2).
FT   CONFLICT     354    355       QD -> KT (IN REF. 2).
FT   CONFLICT     549    549       T -> K (IN REF. 2).
FT   CONFLICT     572    572       T -> R (IN REF. 2).
FT   CONFLICT     575    575       D -> V (IN REF. 2).
FT   CONFLICT     589    594       MQPSES -> SAATEF (IN REF. 2).
FT   CONFLICT     600    602       PEL -> REP (IN REF. 2).
FT   CONFLICT     657    657       H -> D (IN REF. 2 AND 3).
SQ   SEQUENCE   725 AA;  80296 MW;  C2AEB8B4461C6B2F CRC64;


  Query Match             12.3%;  Score 159;  DB 1;  Length 725;
  Best Local Similarity   27.5%;  Pred. No. 1.4e-05;
  Matches   57;  Conservative   35;  Mismatches   81;  Indels   34;  Gaps    9;

Qy      41 QASLIC--------------AVQNHTPEEELLWYREEGPVDLKSGNKINSSSVCVSSISE 86
           | :| |                : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTLITCEASGDPIPSITWRTSTRNISSEEQDL----DGHMVVRSHARV--SSLTLKSIQY 378

Qy      87 NDNGISFIC----RLGRDQSVSVSVVILNVIFPPLLSGNDFQIVEEGSNVKLVCNVKANPQ 142
             | |   : |     :|:|   | |: |   : | | |        ||: | : | | | |
Db     379 RDAG-EYMCIASNPIGQD---SQSIDLEFQYAPKIQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     143 AQMMWYKNSSLIDLEK-SRHQIQQTSESFQLSITKVEKPDNGIYSCIAKSSLKTFSLDFH 201
           | : |:::  ||     |  :|  |  :  | :|    : | | |:| | :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     202 LIVEDKIVGVP----IEPIIAACVVIF 224
           |:  | |   |     :|| :   | |
Db     495 LVQAD-IPSSPSIDRVEPYSSTAQVQF 520


RESULT   9
NCA1_MOUSE
ID   NCA1_MOUSE      STANDARD;      PRT:  1115 AA.
AC   P13595; Q61949;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, large isoform precursor (N-CAM 180)
DE   (NCAM-180) [Contains: N-CAM 140 (NCAM-140)].
GN   NCAM1 OR NCAM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (N-CAM 180).
RC   STRAIN=C57BL/6;
RX   MEDLINE=87246524; PubMed=3595563;
RA   Barthels D., Santoni M.J., Wille W., Ruppert C., Chaix J.C.,
RA   Hirsch M.R., Fontecilla-Camps J.C., Goridis C.;
RT   "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT   a Mr 79,000 polypeptide without a membrane-spanning region.";
RL   EMBO J. 6:907-914(1987)
RN   [2]
RP   SEQUENCE OF 529-1115 FROM N.A. (N-CAM 140).
RC   STRAIN=C57BL/6;
RX   MEDLINE=88067687; PubMed=3684567;
RA   Santoni M.-J., Barthels D., Barbas J.A., Hirsch M.-R., Steinmetz M.,
RA   Goridis C., Wille W.;
RT   "Analysis of cDNA clones that code for the transmembrane forms of the
RT   mouse neural cell adhesion molecule (NCAM) and are generated by
RT   alternative RNA splicing.";
RL   Nucleic Acids Res. 15:8621-8641(1987).
RN   [3]
RP   SEQUENCE OF 642-1115 FROM N.A. (N-CAM 180).
RX   MEDLINE=88283628; PubMed=3396534;
RA   Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.;
RT   "Differential splicing and alternative polyadenylation generates
RT   distinct NCAM transcripts and proteins in the mouse.";
RL   EMBO J. 7:625-632(1988).
RN   [4]
RP   SEQUENCE OF 804-1081 FROM N.A. (N-CAM 180).
RC   STRAIN=C57BL/6J; TISSUE=Brain;
RX   MEDLINE=88247737; PubMed=2454455;
RA   Barthels D., Vopper G., Wille W.;
RT   "NCAM-180, the large isoform of the neural cell adhesion molecule of
RT   the mouse, is encoded by an alternatively spliced transcript.";
RL   Nucleic Acids Res. 16:4217-4225(1988)
RN   [5]
RP   SEQUENCE OF 702-1115 FROM N.A. (N-CAM 140).
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=89251563; PubMed=2721486;
RA   Santoni M.J., Barthels D., Vopper G., Boned A., Goridis C., Wille M.;
RT   "Differential exon usage involving an unusual splicing mechanism
RT   generates at least eight types of NCAM cDNA in mouse brain.";
RL   EMBO J. 8:385-392(1989).
RN   [6]
RP   SEQUENCE OF 20-36.
RX   MEDLINE=86140120; PubMed=3512556;
RA   Rougon G., Marshak D.R.;
RT   "Structural and immunological characterization of the amino-terminal
RT   domain of mammalian neural cell adhesion molecules.";
RL   J. Biol. Chem. 261:3396-3401(1986).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07200; CAA30177.1; -.
DR   EMBL; Y00051; -; NOT_ANNOTATED_CDS.
DR   EMBL; X06328; CAA29641.1; -.
DR   EMBL; X07195; CAA30173.1; -.
DR   EMBL; X07244; CAA30230.1; -.
DR   EMBL; X15051; CAA33150.1; -.
DR   EMBL; X15052; CAA33151.1; -.
DR   PIR; A29673; IJMSNL.
DR   MGD; MGI:97281; Ncam.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.
FT   SIGNAL        1     19
FT   CHAIN        20   1115       NEURAL CELL ADHESION MOLECULE, LARGE
FT                                 ISOFORM.
FT   DOMAIN       20    711       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    712    729       POTENTIAL.
FT   DOMAIN      730   1115       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    295       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      323    393       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      420    487       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      519    596       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      625    692       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       PROBABLE.
FT   DISULFID    139    189       PROBABLE.
FT   DISULFID    235    288       PROBABLE.
FT   DISULFID    330    386       PROBABLE.
FT   DISULFID    427    480       PROBABLE.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    348    348       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    424    424       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    450    450       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    479    479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    810   1076       MISSING (IN ISOFORM N-CAM 140).
SQ   SEQUENCE   1115 AA;  119351 MW;  2C93DCD474CFBCAF CRC64;


  Query Match             12.3%;  Score 159;  DB 1;  Length 1115;
  Best Local Similarity   27.5%;  Pred. No. 2.4e-05;
  Matches   57;  Conservative   35,  Mismatches   81;  Indels   34;  Gaps    9;

Qy      41 QASLIC--------------AVQNHTKEEELLWYKEEGKVDLKSGNKINSSSVCVSSISE 86
           | :| |               : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEQDL----DGHMVVRSHARV--SSLTLKSIQY 378

Qy      87 NDNGISFFC----RDGRQSVSVSVVLNVTPPPLLSGNDFQTVEDGSNVKLVCNVKANPQ 142
             | |  : |     :|:|   | |: |   : | | |        ||: | : | | | |
Db     379 RDAG-EYMCTASNTIGQD---SQSIDLEFQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     143 AQMMWYKNSSLLDLEK-SKHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 201
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     202 LIVKDKTVGVP----IEPIIAACVVIF 224
           |:  | |   |    :|| :   | |
Db     495 LVQAD-TPSSPSIDRVEPYSSTAQVQF 520


RESULT  10
LAR_DROME
ID   LAR_DROME      STANDARD;      PRT;  2029 AA.
AC   P16621;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Protein-tyrosine phosphatase Lar precursor (EC 3.1.3.48) (Protein-
```

```
DE   tyrosine-phosphate phosphohydrolase) (DLAR).
GN   LAR.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90046860; PubMed=2554325;
RA   Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;
RT   "A family of receptor-linked protein tyrosine phosphatases in humans
RT   and Drosophila.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=96178473; PubMed=8598047;
RA   Krueger N.X., van Vactor D., Wan H.I., Gelbart W.M., Goodman C.S.,
RA   Saito H.;
RT   "The transmembrane tyrosine phosphatase DLAR controls motor axon
RT   guidance in Drosophila.";
RL   Cell 84:611-622(1996).
CC   -!- FUNCTION: IT IS POSSIBLE THAT DLAR IS A CELL ADHESION RECEPTOR.
CC       IT POSSESSES AN INTRINSIC PROTEIN TYROSINE PHOSPHATASE ACTIVITY
CC       (PTPase). IT CONTROLS MOTOR AXON GUIDANCE.
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND
CC       PIONEER NEURONS IN THE EMBRYO.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 16 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27700; AAA28668.1; -.
DR   EMBL; U36857; AAC47002.1; -.
DR   EMBL; U36849; AAC47002.1; JOINED.
DR   EMBL; U36850; AAC47002.1; JOINED.
DR   EMBL; U36851; AAC47002.1; JOINED.
DR   EMBL; U36852; AAC47002.1; JOINED.
DR   EMBL; U36853; AAC47002.1; JOINED.
DR   EMBL; U36854; AAC47002.1; JOINED.
DR   EMBL; U36855; AAC47002.1; JOINED.
DR   EMBL; U36856; AAC47002.1; JOINED.
DR   PIR; A36182; TDFFLK.
DR   HSSP; P28827; 1RPM.
DR   FlyBase; FBgn0000464; Lar.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; PnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_prot_phphtase.
DR   Pfam; PF00041; fn3; 9.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00102; Y_phosphatase; 2.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00060; FN3; 9.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00194; PTPc; 2.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW   Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
KW   Cell adhesion; Immunoglobulin domain; Repeat.
FT   SIGNAL       1     32
FT   CHAIN       33   2029       PROTEIN-TYROSINE PHOSPHATASE LAR.
FT   DOMAIN      33   1377       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1378   1402       POTENTIAL.
FT   DOMAIN    1403   2029       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      50    118       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN     154    216       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     249    308       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN     320    417       FIBRONECTIN TYPE-III 1.
FT   DOMAIN     418    512       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     513    607       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     608    706       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     707    809       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     810    906       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     907   1007       FIBRONECTIN TYPE-III 7.
FT   DOMAIN    1008   1102       FIBRONECTIN TYPE-III 8.
FT   DOMAIN    1103   1207       FIBRONECTIN TYPE-III 9.
FT   DOMAIN    1492   1738       PROTEIN-TYROSINE PHOSPHATASE 1.
FT   DOMAIN    1781   2029       PROTEIN-TYROSINE PHOSPHATASE 2.
FT   ACT_SITE  1670   1670       BY SIMILARITY.
FT   ACT_SITE  1961   1961       BY SIMILARITY.
FT   DISULFID    57    111       POTENTIAL.
FT   DISULFID   161    209       POTENTIAL.
FT   DISULFID   256    301       POTENTIAL.
FT   CARBOHYD   176    176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   721    721       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   774    774       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   915    915       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   962    962       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1183   1183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1304   1304       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2029 AA;  229027 MW;  536A0C794D3DC800 CRC64;
```

```
  Query Match             11.8%;  Score 153;  DB 1;  Length 2029;
  Best Local Similarity   25.9%;  Pred. No. 0.00016;
  Matches   51;  Conservative   29;  Mismatches   77;  Indels   40;  Gaps    9;

Qy      39 GSQASLICAVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISENDNGISF----T 94
           |  ||  || :        ::|         |:| |: | :   :: |   |||
Db      50 GGVASFYCAARGDP-PPSIVW--------RKNGKKV-SGTQSRYTVLEQPGGISILRIEP 99

Qy      95 CRLGRD------------QSVSVSVVLNV--------TFPPLLSGNDFQTVEEGSNVKL 133
           | |||             :||    | :          || :  |   : :| |  | :
Db     100 VRAGRDDAPYECVAENGVGDAVSADATLTIYEGDKTPAGFPVITQGPGTRVIEVGHTVLM 159

Qy     134 VCNVKANPQAQMMWYKNSSLLDLEKSKHQTQQISESFQLSTFKVEKPDNGTYSCIAKSSL 193
           |    ||   : :| || : :|:   |: ::   : | | |    : | | | |:|::|:
Db     160 TCKAIGNPTPNIYWIKNQFKVDMSNPRYSLK---DGF-LQIENSREEDQGKYECVAENSM 215

Qy     194 KTE-SLDFHLIVKDKTV 209
           || |   :| || : |
Db     216 GTEHSKATNLYVKVRRV 232
```

```
RESULT  11
DCC_HUMAN
ID   DCC_HUMAN      STANDARD;      PRT;  1447 AA.
AC   P43146;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).
GN   DCC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95011532; PubMed=7926722;
RA   Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA   Vogelstein B.;
RT   "The DCC gene product in cellular differentiation and colorectal
RT   tumorigenesis.";
RL   Genes Dev. 8:1174-1183(1994).
RN   [2]
RP   SEQUENCE OF 1-750 FROM N.A.
RX   MEDLINE=90100559; PubMed=2294591;
RA   Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA   Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA   Vogelstein B.;
RT   "Identification of a chromosome 18q gene that is altered in
RT   colorectal cancers.";
RL   Science 247:49-56(1990).
RN   [3]
RP   SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).
RX   MEDLINE=91121517; PubMed=1991322;
RA   Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
RA   Oliner J.D., Kinzler K.W., Vogelstein B.;
RT   "Scrambled exons.";
RL   Cell 64:607-613(1991).
RN   [4]
RP   GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375
RX   MEDLINE=94245241; PubMed=8188295;
RA   Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
RA   Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
RT   "The DCC gene: structural analysis and mutations in colorectal
RT   carcinomas.";
RL   Genomics 19:525-531(1994).
RN   [5]
RP   VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX   MEDLINE=94243823; PubMed=8187090;
RA   Miyake S., Nagai K., Yoshino K., Uto M., Endo M., Yuasa Y.;
RT   "Point mutations and allelic deletion of tumor suppressor gene DCC in
RT   human esophageal squamous cell carcinomas and their relation to
RT   metastasis.";
RL   Cancer Res. 54:3007-3010(1994).
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC       NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE
CC   -!- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
CC       DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
CC       EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC       POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC       METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X76132; CAA53735.1; -.
DR   EMBL; M32292; AAA35751.1; -.
DR   EMBL; M32286; AAA52174.1; -.
DR   EMBL; M32288; AAA52175.1; ALT_SEQ.
DR   EMBL; M32290; AAA52176.1; -.
DR   EMBL; M63696; AAA52177.1; -.
DR   EMBL; M63700; AAA52178.1; -.
DR   EMBL; M63702; AAA52179.1; -.
DR   EMBL; M63718; AAA52180.1; -.
DR   EMBL; M63698; AAA52181.1; -.
DR   PIR; A54100; A54100.
DR   PIR; A40098; A40098.
DR   PIR; A38442; A38442.
DR   HSSP; P56276; 1TLK.
DR   MIM; 120470; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Glycoprotein, Immunoglobulin domain, Transmembrane; Signal;
KW   Repeat, Anti-oncogene, Disease mutation, Polymorphism.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26   1447       TUMOR SUPPRESSOR PROTEIN DCC.
FT   DOMAIN      26   1097       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1098   1122       POTENTIAL.
FT   DOMAIN    1123   1447       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      54    124       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN     154    219       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     254    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN     345    407       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN     426    522       FIBRONECTIN TYPE-III 1.
FT   DOMAIN     525    618       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     619    716       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     722    816       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     840    940       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     941   1042       FIBRONECTIN TYPE-III 6.
FT   DISULFID    61    117       BY SIMILARITY.
FT   DISULFID   161    212       BY SIMILARITY.
FT   DISULFID   261    310       BY SIMILARITY.
FT   DISULFID   352    400       BY SIMILARITY.
FT   CARBOHYD    94     94       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   299    299       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   318    318       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   478    478       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   628    628       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   702    702       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT    168    168       M -> T (IN OESOPHAGEAL CARCINOMA).
FT                               /FTId=VAR_003909.
FT   VARIANT    201    201       R -> G.
FT                               /FTId=VAR_003910.
FT   VARIANT   1375   1375       P -> H (IN A COLORECTAL CARCINOMA).
FT                               /FTId=VAR_003911.
FT   CONFLICT   138    138       MISSING (IN REF. 3).
FT   CONFLICT   233    329       MISSING (IN REF. 3).
FT   CONFLICT   421    421       MISSING (IN REF. 3).
SQ   SEQUENCE   1447 AA;  158456 MW;  4A8612766ED0471F CRC64;


  Query Match             11.5%;  Score 149;  DB 1;  Length 1447;
  Best Local Similarity   25.8%;  Pred. No. 0.00022;
  Matches   51;  Conservative   37;  Mismatches   86;  Indels   24;  Gaps    9;

Qy      11 ILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKS 70
           : || |   |:|: : ||             | | | |  :       | | |  : |:|
Db     240 LYFLQR--PSNVVAIEGK-----------DAVLECCVSGYP-PPSFTWLRGEEVIQLRS 284

Qy      71 G--NKINSSSVCVSSISENDNGISFTCRLG-RDQSVSVSVVLNVTFPP-LLSGNDFQTVE 126
           : :  |:: :|:::::|:|: :|| :  :::::| | | | ||  |:
Db     285 KKYSLLGGSNLLISNVTDDDSGM-YTCVVTYKNENISASAELTVLVPPWFLNHPSNLYAY 343

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYS 186
           |  :::  | |   |  : | ||  :: :     ||  |    | | | | | | |
Db     344 ESMDIEFECTVSGKPVPTVNWMKNGDVV-IPSDYFQIVGGS---NLRILGVVKSDEGFYQ 399

Qy     187 CIAKSSLKTESLDFHLIV 204
```

```
          |:|::          |||
Db    400 CVAENEAGNAQTSAQLIV 417


RESULT  12
PTPD_HUMAN
ID   PTPD_HUMAN     STANDARD;      PRT;  1912 AA.
AC   P23468;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Protein-tyrosine phosphatase delta precursor (EC 3.1.3.48) (R-PTP-
DE   delta).
GN   PTPRD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.
RX   MEDLINE=95204468; PubMed=7896816;
RA   Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;
RT   "Molecular characterization of the human transmembrane protein-
RT   tyrosine phosphatase delta. Evidence for tissue-specific expression of
RT   alternative human transmembrane protein-tyrosine phosphatase delta
RT   isoforms.";
RL   J. Biol. Chem. 270:6722-6728(1995).
RN   [2]
RP   SEQUENCE OF 690-1912 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=91006018; PubMed=2170109;
RA   Krueger N.X., Streuli M., Saito H.;
RT   "Structural diversity and evolution of human receptor-like protein
RT   tyrosine phosphatases.";
RL   EMBO J. 9:3241-3252(1990).
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT
CC       TISSUES DUE TO ALTERNATIVE SPLICING.
CC   -!- PTM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN
CC       FROM THE TRANSMEMBRANE SEGMENT.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L38929; AAC41749.1; -.
DR   EMBL; X54133; CAA38068.1; -.
DR   PIR; S12052; S12052.
DR   HSSP; P18052; 1YFO.
DR   MIM; 601598; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_prot_phphtase.
DR   Pfam; PF00041; fn3; 8.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00102; Y_phosphatase; 2.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00060; FN3; 8.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00194; PTPc; 2.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW   Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   1912       PROTEIN-TYROSINE PHOSPHATASE DELTA.
FT   DOMAIN      21   1265       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1266   1290       POTENTIAL.
FT   DOMAIN    1291   1912       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      23    115       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN     118    225       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     232    318       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN     320    414       FIBRONECTIN TYPE-III 1.
FT   DOMAIN     417    513       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     516    606       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     609    708       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     711    822       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     825    916       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     918   1017       FIBRONECTIN TYPE-III 7.
FT   DOMAIN    1020   1137       FIBRONECTIN TYPE-III 8.
FT   DOMAIN    1375   1618       PROTEIN-TYROSINE PHOSPHATASE 1.
FT   DOMAIN    1619   1912       PROTEIN-TYROSINE PHOSPHATASE 2.
FT   ACT_SITE  1553   1553       BY SIMILARITY.
FT   ACT_SITE  1844   1844       BY SIMILARITY.
FT   SITE      1175   1178       CLEAVAGE (POTENTIAL).
FT   CARBOHYD   254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   724    724       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   832    832       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   181    189       MISSING (IN KIDNEY ISOFORM).
FT   VARSPLIC   226    229       MISSING (IN KIDNEY ISOFORM).
FT   VARSPLIC   775    783       MISSING (IN KIDNEY ISOFORM).
FT   VARSPLIC   609   1137       MISSING (IN FETAL BRAIN ISOFORM).
FT   MUTAGEN   1178   1178       R->A: 2.5-FOLD REDUCTION IN CLEAVAGE.
SQ   SEQUENCE   1912 AA,  214759 MW,  3AE8CBCB32182F26 CRC64;


  Query Match             11.5%;  Score 149;  DB 1;  Length 1912;
  Best Local Similarity   23.5%;  Pred. No. 0.00032;
  Matches   58;  Conservative   37;  Mismatches   90;  Indels   62;  Gaps   11;

Qy       1 MQMGPFLLLVILFL-------PPEMTSSVLTVNGKTFNYILDTTPGSQASLICAVQNHTR 53
           : : | |||:: |         |   | : :   |           | || ||      |
Db       2 VHVARLLLLLLIFFLRIDAEIPPEFTRTPVDQIG---------VSGGVASFICQATGDPR 52

Qy      54 EEELLWYREEGRVDLESCNKTNSSSVCVSSTSENDNGISFTCPL-----GPDQS------ 102
             :::|         | | |:::       : | |:|     |:       ||::
Db      53 -PKIVWN--------KKGKKVSNQRF---EVIEFDDGSGSVLRIQPLRTPRDEAIYECVA 100

Qy     103 ------VSVSVVLNVT--------FPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWY 148
                :|||  | |           || :  |   : ||       ::|    ||  :: |:
Db     101 SNNVGEISVSFRLIVLREDQIPRGFPIDMGPQLKVVERTRTATMLCAASGNPDPEITWF 160

Qy     149 KNSSLLDLEKSRHQTQQ-ISESF-------QLSTTKVEKPDNGTYSCIAKSSLKTE-SLD 199
           |:   :|   :  :|:|  |||         | | : |: | | | |:| :|  |  |
Db     161 KDFLPVDTSNNNGRIKQLRSESTQGTPTRGALQIEQSEESDQGKYECVAINSAGIRYSAP 220

Qy     200 FHLIVKD 206
             :| |::
Db     221 ANLYVRE 227


RESULT  13
LAMP_HUMAN
ID   LAMP_HUMAN     STANDARD;      PRT;   338 AA.
AC   Q13449;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Limbic system-associated membrane protein precursor (LSAMP).
GN   LSAMP OR LAMP.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96235133; PubMed=8666243;
RA   Pimenta A.F., Fischer I., Levitt P.;
RT   "cDNA cloning and structural analysis of the human limbic-system-
RT   associated membrane protein (LAMP).";
RL   Gene 170:189-195(1996).
CC   -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
CC       CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC       MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC       OF THE HYPPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON LIMBIC NEURONS AND FIBER TRACTS
CC       AS WELL AS IN SINGLE LAYERS OF THE SUPERIOR COLLICULUS, SPINAL
CC       CHORD AND CEREBELLUM.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U41901; AAC50569.1; -.
DR   MIM; 603241; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2
KW   Immunoglobulin domain, Cell adhesion, Glycoprotein, GPI anchor,
KW   Repeat; Signal.
FT   SIGNAL       1     28       POTENTIAL.
FT   CHAIN       29    315       LIMBIC SYSTEM-ASSOCIATED MEMBRANE
FT                               PROTEIN.
FT   PROPEP     316    338       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN      46    118       IG-LIKE C2-TYPE DOMAIN 1
FT   DOMAIN     146    204       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     232    297       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID    53    111       POTENTIAL.
FT   DISULFID   153    197       POTENTIAL.
FT   DISULFID   239    290       POTENTIAL.
FT   CARBOHYD    40     40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    66     66       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   300    300       N-LINKED (GLCNAC     ) (POTENTIAL).
FT   CARBOHYD   315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID      315    315       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   338 AA;  37308 MW;  03455F286DF5D92F CRC64;


  Query Match           11.5%;  Score 148.5;  DB 1;  Length 338;
  Best Local Similarity  25.4%;  Pred. No. 3.8e-05;
  Matches   60;  Conservative   34;  Mismatches   97;  Indels   45;  Gaps   12;

Qy       2 QMGRFLLLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR 61
           |:  || ::  ||  :   -  |  |:|  -    |  | | | ···  -  ·· |
Db      11 QLPLVLLRLLCLLPTGLPVRSVDFNRGTDN--ITVRQGDTAILRCVLED--KNSKVAWLN 66

Qy      62 EEG------------RVDLKSGNKINSSSVCVSSISENDNGISFTCRLGRDQSVSVS-V 107
             |             ||:|:  : :   |: :  :   | | |:|| :        | |
Db      67 RSGIIFAGHDKWSLDPRVELEKRHSL-EYSLRIQKVDVYDEG-SYTCSVQTQHEPKTSQV 124

Qy     108 VLNVTFPPLLSG-NDPQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQT 166
           | |  || :|  :   || ||||| |||     |:  : |            || :  |
Db     125 YLIVQVPPKISNISSDVTVNEGSNVTLVCMANGRPEPVITW-----------RH-LTPT 171

Qy     167 SESFQ--- -LSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGVPIEPII 217
                 |:     | |  : :  :| | |  |:: :  | |    ||   | |   | |
Db     172 GREPEGEEEYLEILGITREQSGKYEC--KAANEVSSAD----VKQVKVTVNYPPTI 221


RESULT  14
PGBM_HUMAN
ID   PGBM_HUMAN     STANDARD;      PRT;  4393 AA.
AC   P98160; Q16287;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin, and Colon;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 1018-1472 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [4]
RP   SEQUENCE OF 892-1398 FROM N.A.
RC   TISSUE=Fibrosarcoma;
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [5]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
CC   -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC       MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC       CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
```

```
CC       PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC       SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC       SUBSTRATE FOR CELLS.
CC     !- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC       DIMERS OR STELLATE STRUCTURES.
CC     -!- SUBCELLULAR LOCATION: Extracellular.
CC     -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC     -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC     -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC     -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC     -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC     -!- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC     -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC     -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC     -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC     ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   HSSP; P01130; 1AJJ.
DR   Siena-2DPAGE; P98160; -.
DR   MIM; 142461; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_CA.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 8.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00001; EGF_like; 8.
DR   SMART; SM00408; IGc2; 22.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4393       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      197    236       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      283    321       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      323    361       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      366    405       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      405    506       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      523    532       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      533    732       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      733    765       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      766    815       LAMININ EGF-LIKE 2.
FT   DOMAIN      816    873       LAMININ EGF-LIKE 3.
FT   DOMAIN      881    925       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      926    935       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      936   1127       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1128   1160       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1161   1210       LAMININ EGF-LIKE 6.
FT   DOMAIN     1211   1267       LAMININ EGF-LIKE 7.
FT   DOMAIN     1277   1326       LAMININ EGF-LIKE 8.
FT   DOMAIN     1327   1336       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1337   1531       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1532   1564       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1565   1614       LAMININ EGF-LIKE 10.
FT   DOMAIN     1615   1672       LAMININ EGF-LIKE 11.
FT   DOMAIN     1679   1773       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     1774   1867       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN     1868   1957       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN     1958   2053       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN     2054   2153       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN     2154   2246       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN     2247   2342       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN     2343   2438       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN     2439   2535       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     2536   2631       IG-LIKE C2-TYPE DOMAIN 11.
FT   DOMAIN     2632   2728       IG-LIKE C2-TYPE DOMAIN 12.
FT   DOMAIN     2729   2828       IG-LIKE C2-TYPE DOMAIN 13.
FT   DOMAIN     2829   2926       IG-LIKE C2-TYPE DOMAIN 14.
FT   DOMAIN     2927   3023       IG-LIKE C2-TYPE DOMAIN 15.
FT   DOMAIN     3024   3114       IG-LIKE C2-TYPE DOMAIN 16.
FT   DOMAIN     3115   3213       IG-LIKE C2-TYPE DOMAIN 17.
FT   DOMAIN     3214   3300       IG-LIKE C2-TYPE DOMAIN 18.
FT   DOMAIN     3301   3401       IG-LIKE C2-TYPE DOMAIN 19.
FT   DOMAIN     3402   3490       IG-LIKE C2-TYPE DOMAIN 20.
FT   DOMAIN     3491   3576       IG-LIKE C2-TYPE DOMAIN 21.
FT   DOMAIN     3577   3664       IG-LIKE C2-TYPE DOMAIN 22.
FT   DOMAIN     3665   3845       LAMININ G-LIKE 1.
FT   DOMAIN     3846   3883       EGF-LIKE 1.
FT   DOMAIN     3886   3924       EGF-LIKE 2.
FT   DOMAIN     3930   4105       LAMININ G-LIKE 2.
FT   DOMAIN     4106   4143       EGF-LIKE 3.
FT   DOMAIN     4145   4178       EGF-LIKE 4.
FT   DOMAIN     4203   4391       LAMININ G-LIKE 3.
FT   SITE         65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE       4151   4153       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   SITE       4301   4303       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    206    225       BY SIMILARITY.
FT   DISULFID    219    234       BY SIMILARITY.
FT   DISULFID    285    297       BY SIMILARITY.
FT   DISULFID    292    310       BY SIMILARITY.
FT   DISULFID    304    319       BY SIMILARITY.
FT   DISULFID    325    337       BY SIMILARITY.
FT   DISULFID    342    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.

  Query Match             11.4%;  Score 147.5;  DB 1;  Length 4393;
  Best Local Similarity;  26.2%,  Prob. No.. 0.0012;
  Matches   55;  Conservative   38;  Mismatches   78;  Indels   39;  Gaps   12;

Qy      19 TSSVLEVNGNTENYIIDITPGSQASLDCAVGNHTPEEELLWYPEEGPVDLKSGNKINSSS 78
             :||    || |    ||         |: :   ||    : ||:  |   | | ::|  |
```

```
Db   2542 SSSASLANGHT----LDLN-----CLVASQAPHT----ITWYKRGG--SLPSRHQIVGSR 2586

Qy     79 VCVSSISENDNGISFTCRL-----GRDQSVSVSV-------VLNVTFPPLLSGNDFQTVE 126
          : :  ::  |:|  : | :      |: |: |::       |  |: ||:   :   ||
Db   2587 LRIPQVTPADSG-EYVCHVSNGAGSRETSLIVTIQGSGSSHVPRVS-PPIRIESSSPTVV 2644

Qy    127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYS 186
          ||  : | | |   ||| : |||    |    ||||    :    | : ::   |:| |
Db   2645 EGQTLDLNCVVARQPQAIITWYKRGGSL---PSRHQ----THGSHLRLHQMSVADSGEYV 2697

Qy    187 CIAKSSLKTESLDFHLIVK-DKTVGVPIEP 215
          | | :::  ::|:  :::    : | |  |
Db   2698 CRANNNI--DALEASIVISVSPSAGSPSAP 2725


RESULT  15
DCC_MOUSE
ID   DCC_MOUSE      STANDARD;      PRT;  1447 AA.
AC   P70211;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor.
GN   DCC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Brain;
RX   MEDLINE=96112625; PubMed=8570174;
RA   Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F.;
RT   "Cloning of the mouse homologue of the deleted in colorectal cancer
RT   gene (mDCC) and its expression in the developing mouse embryo.";
RL   Oncogene 11:2243-2254(1995).
RN   [2]
RP   REVISIONS.
RC   STRAIN=BALB/C; TISSUE=Brain;
RA   Cooper H.M.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM
CC       THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD
CC       FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC       DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC       BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC       EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC       AND REMAIN AT THIS LEVEL IN THE ADULT.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85788; CAA59786.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:94869; Dcc.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW   Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1447       TUMOR SUPPRESSOR PROTEIN DCC, LONG
FT                                ISOFORM.
FT   CHAIN        85   1447       TUMOR SUPPRESSOR PROTEIN DCC, SHORT
FT                                ISOFORM.
FT   INIT_MET     85     85       FOR SHORT ISOFORM.
FT   DOMAIN       26   1097       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1098   1122       POTENTIAL.
FT   DOMAIN     1123   1447       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       54    124       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      154    219       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      254    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      345    407       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      426    522       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      525    618       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      619    716       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      722    816       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      840    940       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      941   1042       FIBRONECTIN TYPE-III 6.
FT   DISULFID     61    117       BY SIMILARITY.
FT   DISULFID    161    212       BY SIMILARITY.
FT   DISULFID    261    310       BY SIMILARITY.
FT   DISULFID    352    400       BY SIMILARITY.
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     94     94       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    628    628       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    702    702       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    819    838       MISSING (IN EMBRYONIC ISOFORM).
SQ   SEQUENCE   1447 AA;  158298 MW;  0D1F1097C22D5B9F CRC64;


  Query Match           11.4%;  Score 147;  DB 1;  Length 1447;
  Best Local Similarity  25.8%;  Pred. No. 0.00032;
  Matches   51;  Conservative   37;  Mismatches   86;  Indels   24;  Gaps    9;

Qy     11 ILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKS 70
          : || |   |:|: : ||           | | | |  :         | | |  : |:|
Db    240 LYFLQR--PSNVIAIEGK------------DAVLECCVSGYP-PPSFTWLRGEEVIQLRS 284

Qy     71 G--NKINSSSVCVSSISENDNGISFTCRLG-RDQSVSVSVVLNVTFPP-LLSGNDFQTVE 126
          : :  |:: :|:::::|:| ::|| :  :::::| |  | |  ||  |:
Db    285 KKYSLLGGSNLLISNVTDDDSG-TYTCVVTYKNENISASAELTVLVPPWFLNHPSNLYAY 343

Qy    127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYS 186
          |  :::  | |   |   : | ||  :: :     ||   |    | |  | | | | |
Db    344 ESMDIEFECAVSGKPVPTVNWMKNGDVV-IPSDYFQIVGGS---NLRILGVVKSDEGFYQ 399

Qy    187 CIAKSSLKTESLDFHLIV 204
          |:|::          |||
Db    400 CVAENEAGNAQSSAQLIV 417


Search completed: September 17, 2002, 13:24:15
Job time: 802 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:24:15 ; Search time 40.16 Seconds
(without alignments)
252.603 Million cell updates/sec

Title: US-09-863-823-6
Perfect score: 1336
Sequence: 1 MAWKSSVIMQMGRFLLLVIL.....RKKIMKLCMKDKDPHSETAL 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	176	13.2	853	1	NCA1_BOVIN	P31836 bos taurus
2	175	13.1	858	1	NCA1_RAT	P13596 rattus norv
3	173	12.9	359	1	LACH_DROME	Q24372 drosophila
4	170	12.7	2012	1	DSCA_HUMAN	O60469 homo sapien
5	166	12.4	761	1	NCA2_HUMAN	P13592 homo sapien
6	166	12.4	848	1	NCA1_HUMAN	P13591 homo sapien
7	160.5	12.0	1091	1	NCA1_CHICK	P13590 gallus gall
8	159	11.9	725	1	NCA2_MOUSE	P13594 mus musculu
9	159	11.9	1115	1	NCA1_MOUSE	P13595 mus musculu
10	153	11.5	2029	1	LAR_DROME	P16621 drosophila
11	150	11.2	1912	1	PTPD_HUMAN	P23468 homo sapien
12	149	11.2	1447	1	DCC_HUMAN	P43146 homo sapien
13	148.5	11.1	338	1	LAMP_HUMAN	Q13449 homo sapien
14	147.5	11.0	4393	1	PGBM_HUMAN	P98160 homo sapien
15	147	11.0	1447	1	DCC_MOUSE	P70211 mus musculu
16	145	10.9	837	1	NCM2_HUMAN	O15394 homo sapien
17	144.5	10.8	338	1	LAMP_CHICK	Q98919 gallus gall
18	144.5	10.8	338	1	LAMP_RAT	Q62813 rattus norv
19	144.5	10.8	1348	1	VGR2_COTJA	P52583 coturnix co
20	144	10.8	521	1	CEA1_MOUSE	P31809 mus musculu
21	142.5	10.7	349	1	LACH_SCHAM	Q26474 schistocerc
22	142	10.6	837	1	NCM2_MOUSE	O35136 mus musculu
23	142	10.6	1356	1	VGR2_HUMAN	P35968 homo sapien
24	141.5	10.6	1493	1	NEO1_MOUSE	P97798 mus musculu
25	139	10.4	1450	1	MPSF_CHICK	Q02173 gallus gall
26	137	10.3	1092	1	NCA2_XENLA	P36335 xenopus lae
27	137	10.3	1343	1	VGR2_RAT	O08775 rattus norv
28	136.5	10.2	519	1	ECTO_RAT	P16573 rattus norv
29	136.5	10.2	1897	1	PTPF_HUMAN	P10586 homo sapien
30	135.5	10.1	1070	1	PTK7_HUMAN	Q13308 homo sapien
31	135	10.1	1377	1	NEO1_RAT	P97603 rattus norv
32	134	10.0	344	1	CEA6_HUMAN	P40199 homo sapien
33	133.5	10.0	3707	1	PGBM_MOUSE	Q05793 mus musculu
34	132.5	9.9	976	1	KIT_HUMAN	P10721 homo sapien
35	131.5	9.8	353	1	CEPU_CHICK	Q90773 gallus gall
36	131.5	9.8	526	1	CEA1_HUMAN	P13688 homo sapien
37	131	9.8	1088	1	NCA1_XENLA	P16170 xenopus lae
38	130.5	9.8	298	1	JAM2_HUMAN	P57087 homo sapien
39	130.5	9.8	702	1	CEA5_HUMAN	P06731 homo sapien
40	130.5	9.8	864	1	TRKC_RAT	Q03351 rattus norv
41	130	9.7	1367	1	VGR2_MOUSE	P35918 mus musculu
42	129.5	9.7	333	1	AMAL_DROME	P15364 drosophila
43	129.5	9.7	626	1	MAG_MOUSE	P20917 mus musculu
44	129.5	9.7	822	1	FGR1_HUMAN	P11362 homo sapien
45	129.5	9.7	822	1	FGR1_MOUSE	P16092 mus musculu

ALIGNMENTS

RESULT 1
NCA1_BOVIN
ID NCA1_BOVIN STANDARD; PRT; 853 AA.
AC P31836;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE (NCAM-140).
GN NCAM1 OR NCAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain cortex;
RX MEDLINE=89378239; PubMed=2776887;
RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshnyakov M.V.,
RA Petukhova G.V., Rakitina T.V., Feshchenko E.A., Ishchenko K.A.,
RA Mirzoeva S.F., Chernova M.N., Dranytsyna S.M.;
RT "Calmodulin-independent bovine brain adenylate cyclase. Amino acid
RT sequence and nucleotide sequence of the corresponding cDNA.";
RL FEBS Lett. 254:69-73(1989).
RN [2]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401(1986).
RN [3]
RP IDENTIFICATION AS N-CAM.
RX MEDLINE=92111748; PubMed=1765159;
RA Premont R.T.;
RT "A bovine brain cDNA purported to encode calmodulin-insensitive
RT adenylyl cyclase has extensive identity with neural cell adhesion
RT molecules (N-CAMs).";
RL FEBS Lett. 295:230-231(1991).
CC -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A CALMODULIN-
CC INDEPENDENT ADENYLATE CYCLASE.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X16451; CAA34470.1; -.
DR   PIR; A32976; IJBONC.
DR   HSSP; P40189; 1BQU.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20    858       NEURAL CELL ADHESION MOLECULE, 140 KDA
FT                                 ISOFORM.
FT   DOMAIN       20    719       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    720    737       POTENTIAL.
FT   DOMAIN      738    858       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    295       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      321    401       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      428    495       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      527    604       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      633    700       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       BY SIMILARITY.
FT   DISULFID    139    189       BY SIMILARITY.
FT   DISULFID    235    286       BY SIMILARITY.
FT   DISULFID    328    394       BY SIMILARITY.
FT   DISULFID    435    488       BY SIMILARITY.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    314    314       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    346    346       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    458    458       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   853 AA;  93893 MW;  E12FD49231A7A368 CRC64;


  Query Match             13.2%;  Score 176;  DB 1,  Length 853;
  Best Local Similarity   28.4%;  Pred. No. 8.9e-07;
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    6;

Qy      49 QASLIC        --------AVQNHTREFEILWYREEGPVDLKSGNKINS----SSVCVS 90
           | :| |               : :| : ||:  | | | : |     : |    ||: :
Db     323 QVTLTCEASGDPIPSITWRTSTRNISSEEKASWTRPEKQETLDGHMVVRSHARVSSLTLK 382

Qy      91 SISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 146
           ||   | |   : |      :|:|   | |: | | : | | |        ||: | : | |
Db     383 SIQYTDAG-EYVCTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 438

Qy     147 ANPQAQMMWYKNSSLLDEK  SPRQSQGTSESELSITKVEKEDNGTYSCIAKSSLAIES 205
           | | | : |::  ||     |  :|  |  :  | :|   : | | |:| | : :  ||
Db     439 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 498

Qy     206 LDFHLIVKDETVGVP----TEPIIAACVVIF 232
           |:| |:  | |   |    :|| :   | |
Db     499 LEFVLVQAD-TPSSPSIDQVEPYSSTAQVQF 528


RESULT   2
NCAM_RAT
ID   NCAM_RAT       STANDARD;      PRT;   858 AA.
AC   P13596;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE   (NCAM-140).
GN   NCAM1 OR NCAM.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=88059265; PubMed=3680385;
RA   Small S.J., Shull G.E., Santoni M.-J., Akeson R.;
RT   "Identification of a cDNA clone that contains the complete coding
RT   sequence for a 140 kD rat NCAM polypeptide.";
RL   J. Cell Biol. 105:2335-2345(1987).
RN   [2]
RP   SEQUENCE OF 355-364 FROM N.A.
RX   MEDLINE=90166485; PubMed=2483093;
RA   Small S.J., Haines S.L., Akeson R.A.;
RT   "Polypeptide variation in an N-CAM extracellular immunoglobulin-like
RT   fold is developmentally regulated through alternative splicing.";
RL   Neuron 1:1007-1017(1988).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC    !  SUBCELLULAR LOCATION. Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X06564; CAA29809.1; -.
DR   EMBL; M32611; AAA41679.1; -.
DR   PIR; S00846; IJRTNC.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.
FT   SIGNAL        1     19
FT   CHAIN        20    858       NEURAL CELL ADHESION MOLECULE, 140 KDA
FT                                 ISOFORM.
FT   DOMAIN       20    721       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    722    739       POTENTIAL.
FT   DOMAIN      740    858       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    295       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      323    403       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      430    497       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      514    615       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      616    712       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       BY SIMILARITY.
FT   DISULFID    139    189       BY SIMILARITY.
FT   DISULFID    235    288       BY SIMILARITY.
FT   DISULFID    330    396       BY SIMILARITY.
FT   DISULFID    437    490       BY SIMILARITY.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    348    348       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    434    434       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    460    460       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   858 AA;  94658 MW;  EA1A06A4EA0550F6 CRC64;


  Query Match             13.1%;  Score 175;  DB 1;  Length 858;
  Best Local Similarity   28.4%;  Pred. No. 1.1e-06;
  Matches   60;  Conservative   31;  Mismatches   88;  Indels   32;  Gaps    8;

Qy      49 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINS----SSVCVS 90
           | :| |               : :| : ||:  | | | :  |     : |    ||: :
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEKASWTRPEKQETLDGHMVVRSHARVSSLTLK 384

Qy      91 SISENDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVK 146
           ||   | |  : |     :|:|   | |: | | : | | |       ||: | : | |
Db     385 SIQYTDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVF 440

Qy     147 ANPQAQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTES 205
           | | | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  ||
Db     441 AYPSATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQES 500

Qy     206 LDFHLIVKDKTVGVP----IEPIIAACVVIF 232
           |:| |:  | |   |    :|| :  | |
Db     501 LEFILVQAD-TPSSPSIDRVEPYSSTAQVQF 530


RESULT   3
LACH_DROME
ID   LACH_DROME     STANDARD;      PRT;   359 AA.
AC   Q24372;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Lachesin precursor.
GN   LAC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94038693; PubMed=8223276;
RA   Karlstrom R.O., Wilder L.P., Bastiani M.J.;
RT   "Lachesin: an immunoglobulin superfamily protein whose expression
RT   correlates with neurogenesis in grasshopper embryos.";
RL   Development 118:509-522(1993).
CC   -!- FUNCTION: MAY PLAY A ROLE IN EARLY NEURONAL DIFFERENTIATION AND
CC       AXON OUTGROWTH.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ON DIFFERENTIATING NEURONAL CELLS
CC       FROM THE ONSET OF NEUROGENESIS IN BOTH THE CENTRAL AND PERIPHERAL
CC       NERVOUS SYSTEMS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L13255; AAC37184.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0010238; Lac.
DR   InterPro, IPR003006, Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain, Cell adhesion, Glycoprotein, GPI anchor;
KW   Repeat; Signal.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    336       LACHESIN.
FT   PROPEP      337    359       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       43    120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      150    211       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      240    310       IG-LIKE C2-TYPE DOMAIN 2.
FT   DISULFID     50    113       POTENTIAL.
FT   DISULFID    157    204       POTENTIAL.
FT   DISULFID    247    303       POTENTIAL.
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       336    336       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   359 AA;  39998 MW;  D1F513E2B9D886E6 CRC64;


  Query Match             12.9%;  Score 173;  DB 1;  Length 359;
  Best Local Similarity   27.3%;  Pred. No. 5.2e-07;
  Matches   60;  Conservative   43;  Mismatches   85;  Indels   32;  Gaps   11;

Qy       4 WKSSVIMQMGRFLLLVILFLPPEMTSSVLIVNGKPENYILLPTPGSQASLICAVQNHTPE 63
           |: |:   :   ||| | |: : :     |::  |:  | |   |       :|| : :|
Db       2 WRPSISNCVWSILLLAI-FVQQILAQKIPITISYTIQPQTNVT---GGIVEPDGSVQ-YAKE 57

Qy      64 EELLWYR-EEGRVDLKSGNK1-------------NSSS--VCVSSISENDNGISFTCR-- 104
             :|: : :   | | :|: :               |||:  : :  | | | | ::||:
Db      58 YNVLFLKTDSDPVFLSTGSTLVIKDSRFSLRYDPNSSTYKLQIKDIQETDAG-TYTCQVV 116

Qy     105 LGRDQSVSVSVVLNVTFPPLLSGNDFQTV--EEGSNVKLVCNVKANPQAQMMWYK-NSSL 161
           :     ||  | |:|  ||::| |  |:|   ||| |:: |     |   : | : |:::
Db     117 ISIVHKVSAEVKLSVKRPPVISDNSIQSVVASEGSEVQMECYASGYPTPTITWRRENNAI 176

Qy     162 LDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSL 201
           |  : :      |     | |  |:| | ||| |:| : :
Db     177 LPTDSA------TYVGNTLRIKSVKKEDRGTYYCVADNGV 210


RESULT   4
DSCA_HUMAN
ID   DSCA_HUMAN     STANDARD;      PRT;  2012 AA.
AC   O60469; O60468;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Down syndrome cell adhesion molecule precursor (CHD2).
GN   DSCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=98087574; PubMed=9426258;
RA   Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA   Lyons G.E., Korenberg J.R.;
RT   "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT   Down syndrome region and is involved in the development of the
RT   nervous system.";
RL   Hum. Mol. Genet. 7:227-237(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
```

```
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=20384934; PubMed=10925149;
RA   Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT   "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT   intercellular adhesion.";
RL   Brain Res. Mol. Brain Res. 79:118-126(2000).
CC   -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC       INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC       NERVOUS SYSTEM DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC       SHORT ISOFORM MAY BE SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/CHD2-52 (SHOWN HERE)
CC       AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF023450; AAC17967.1; -.
DR   EMBL; AF023449; AAC17966.1; -.
DR   EMBL; AL163283; CAB90464.1; -.
DR   EMBL; AL163282; CAB90436.1; -.
DR   EMBL; AL163281; CAB90444.1; -.
DR   HSSP; P40189; 1BQU.
DR   MIM; 602523; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 9.
DR   PRINTS; PR00014; ENTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 7.
KW   Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW   Transmembrane; Alternative splicing.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   2012       DOWN SYNDROME CELL ADHESION MOLECULE.
FT   DOMAIN       18   1595       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1596   1616       POTENTIAL.
FT   DOMAIN     1617   2012       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39    109       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      138    204       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      239    300       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      328    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      421    491       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    582       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      610    676       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      704    773       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      802    872       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN      885    972       FIBRONECTIN TYPE III 1.
FT   DOMAIN      984   1076       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     1088   1177       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     1189   1273       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     1300   1366       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     1380   1463       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     1477   1562       FIBRONECTIN TYPE-III 6.
FT   DISULFID     46    102       BY SIMILARITY.
FT   DISULFID    145    197       BY SIMILARITY.
FT   DISULFID    246    293       BY SIMILARITY.
FT   DISULFID    335    385       BY SIMILARITY.
FT   DISULFID    428    484       BY SIMILARITY.
FT   DISULFID    525    575       BY SIMILARITY.
FT   DISULFID    617    669       BY SIMILARITY.
FT   DISULFID    711    766       BY SIMILARITY.
FT   DISULFID    809    865       BY SIMILARITY.
FT   DISULFID   1307   1359       BY SIMILARITY.
FT   CARBOHYD     28     28       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     78     78       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    512    512       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    748    748       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    795    795       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1142   1142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1160   1160       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1250   1250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1271   1271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1341   1341       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1488   1488       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1562   1571       NFATINYDQS -> KEAARCKEFS (IN SHORT
FT                                ISOFORM).
FT   VARSPLIC   1572   2012       MISSING (IN SHORT ISOFORM).
FT   CONFLICT   1893   2012       HPPGDLIHLPPYIRMDPLLNRGGPGTSRDLSLGQACLFPQK
FT                                SRTLERPTVLEPIPMEAASSASSTREGQSWQPGAVATLPQR
FT                                FGAFLGQAAKMSSSQESLIDSRGHLKGNNFYAKSYTLV ->
FT                                IGQVTSYICLHTLEWTFC (IN REF. 1).
SQ   SEQUENCE   2012 AA;  222259 MW;  0E33CFB781A08334 CRC64;
```

```
  Query Match             12.7%;  Score 170;  DB 1;  Length 2012;
  Best Local Similarity   29.8%;  Pred. No. 8.2e-06;
  Matches   57;  Conservative   31;  Mismatches   69;  Indels   34;  Gaps   10;

Qy      23 PREMISSVLIVNGKIENYTLDFTPGSQASLICAVQNHTREEELLWYREEGPVLESGNK- 81
           ||:: |||                 ||| || |:|   | ::|| |||  | : |  |
Db     320 PRKVKSSV----------------GSQVSLSCSVTG-TEDQELSWYR-NGEI-LNPGKNV 360

Qy      82 ----INSSSVCVSSISENDNGISFTCKLGRDQSVS---VSVVLNVTFPPLLSGNDFQTVE 134
               ||  :: :  : ::| | :: | : :|:  :   | |||    | ::|    : |
Db     361 RITGINHENLIMDHMVKSDGG-AYQCFVRKDKLSAQDYVQVVLEDGTPKIISAFSEKVVS 419

Qy     135 EGSNVKLVCNVEANPQAQMMWYKNSSLLDLEKSRHQIQQ---TSES---FQLSITKVEKPD 189
             | |:|||  |  : | :   : |:   |:| |  |||      |:|:  :  |
Db     420 PAEPVSLMCNVKGTPLPTITWILDDDP-LKGGSHKISQMITSEGNVVSYLNISSSQVRD 478

Qy     190 NGTYSCIAKSS 200
           | | | | :|
Db     479 GGVYRCTANNS 489


RESULT   5
NCA2_HUMAN
ID   NCA2_HUMAN     STANDARD;      PRT;   761 AA.
AC   P13592; P13593;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, phosphatidylinositol linked isoform
DE   precursor (N-CAM 120) (NCAM-120) (CD56 antigen).
```

```
GN   NCAM1 OR NCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (N-CAM 120).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=89305258; PubMed=3253057;
RA   Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,
RA   Elsom V., Moore S.E., Goridis C., Walsh F.S.;
RT   "Complete sequence and in vitro expression of a tissue-specific
RT   phosphatidylinositol-linked N-CAM isoform from skeletal muscle.";
RL   Development 104:165-173(1988).
RN   [2]
RP   SEQUENCE OF 491-761 FROM N.A. (N-CAM 120).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=87301755; PubMed=2887295;
RA   Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA   Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT   "Human muscle neural cell adhesion molecule (N-CAM): identification
RT   of a muscle-specific sequence in the extracellular domain.";
RL   Cell 50:1119-1130(1987).
RN   [3]
RP   SEQUENCE OF 491-655 FROM N.A. (SECRETED ISOFORM).
RX   MEDLINE=89077552; PubMed=3203385;
RA   Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,
RA   Dickson G., Walsh F.S.;
RT   "Alternative splicing generates a secreted form of N-CAM in muscle
RT   and brain.";
RL   Cell 55:955-964(1988).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16841; CAA34739.1; -.
DR   EMBL; M17409; AAA59912.1; -.
DR   EMBL; M22094; AAA59910.1; -.
DR   EMBL; M22092; AAA59911.1; -.
DR   EMBL; M22091; AAA59911.1; JOINED.
DR   PIR; S07784; IJHUNG.
DR   PIR; A31635; A31635.
DR   MIM; 116930; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW   GPI-anchor; Alternative splicing.
FT   SIGNAL        1     19
FT   CHAIN        20    761       NEURAL CELL ADHESION MOLECULE,
FT                                 PHOSPHATIDYLINOSITOL-LINKED ISOFORM.
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    294       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      322    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      419    486       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    595       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      660    727       FIBRONECTIN TYPE-III 2.
FT   DISULFID     41     96       PROBABLE.
FT   DISULFID    139    189       PROBABLE.
FT   DISULFID    235    287       PROBABLE.
FT   DISULFID    329    385       PROBABLE.
FT   DISULFID    426    479       PROBABLE.
FT   CARBOHYD    202    202       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    635    655       GEPSAPKLEGQMGEDGNSIKV -> NIAQNHCCNMFQAGLH
FT                                NALMK (IN SECRETED ISOFORM).
SQ   SEQUENCE   761 AA;  83770 MW;  F0CAD3292D7AB67E CRC64;
```

```
  Query Match             12.4%;  Score 166;  DB 1;  Length 761;
  Best Local Similarity   28.0%;  Pred. No. 5e-06;
  Matches   58;  Conservative   35,  Mismatches   80,  Indels   34;  Gaps    9;

Qy      49 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 94
           | :| |                : :| : ||: |    :| : ::|  ::  ||: : ||
Db     324 QVTLTCEASGDPIPSITWRTSTRNISSEEKTL----DGHMVVRSHARV--SSLTLKSIQY 377

Qy      95 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 150
             | |  : |     :|:|   | |: | | : | | |        ||: | : | | | |
Db     378 IDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYIWEGNQVNITCEVFAYPS 433

Qy     151 AQMMWYKNSSLIDLEK-SPHQIQQTSFSFQLSITKVEKPDNGTYSCIAKSSLKTFSLDFH 209
           | : |:::  ||     |  :| |  : | :|   : | | |:| : | : :  |||:|
Db     434 ATISWFPDGQLLPSSNYSNIKIYNIPSASYLEVIPQSENDFGNYNCTAVNRIGQESLEFI 493

Qy     210 LIVKDKTVGVP----IEPIIAACVVIF 232
           |:  | |   |    :|| :  | |
Db     494 LVQAD-TPSSPSIDQVEPYSSTAQVQF 519
```

```
RESULT   6
NCA1_HUMAN
ID   NCA1_HUMAN     STANDARD;      PRT;   848 AA.
AC   P13591; Q16180; Q15829;
DT   01-JAN-1990 (Rel. 13, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE   (NCAM-140) (CD56 antigen).
GN   NCAM1 OR NCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94356433; PubMed=8075973;
RA   Saito S., Tanio Y., Tachibana I., Hayashi S., Kishimoto T., Kawase I.;
RT   "Complementary DNA sequence encoding the major neural cell adhesion
RT   molecule isoform in a human small cell lung cancer cell line.";
RL   Lung Cancer 10:307-318(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91250739; PubMed=1710251;
RA   Lanier L.L., Chang C., Azuma M., Ruitenberg J.J., Hemperly J.J.,
RA   Phillips J.H.;
RT   "Molecular and functional analysis of human natural killer cell-
RT   associated neural cell adhesion molecule (N-CAM/CD56).";
RL   J. Immunol. 146:4421-4426(1991).
RN   [3]
RP   SEQUENCE OF 491-848 FROM N.A.
```

```
RX   MEDLINE=87401755; PubMed=2887295;
RA   Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA   Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT   "Human muscle neural cell adhesion molecule (N-CAM): identification
RT   of a muscle-specific sequence in the extracellular domain.";
RL   Cell 50:1119-1130(1987).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD quide CD56 entry;
CC       WWW-"http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S71824; AAB31836.1; -.
DR   EMBL; D63041; AAB04558.1; -.
DR   EMBL; M17410; AAA59913.1; -.
DR   PIR; B26883; B26883.
DR   HSSP; P40189; 1BQU.
DR   MIM; 116930; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW   Transmembrane; Alternative splicing.
FT   SIGNAL       1     19
FT   CHAIN       20    848       NEURAL CELL ADHESION MOLECULE, 140 KDA
FT                               ISOFORM.
FT   DOMAIN       20    708       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    709    729       POTENTIAL.
FT   DOMAIN      730    848       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    291       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      322    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      419    486       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    595       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      660    727       FIBRONECTIN TYPE-III 2.
FT   DISULFID     41     96       PROBABLE.
FT   DISULFID    139    189       PROBABLE.
FT   DISULFID    235    287       PROBABLE.
FT   DISULFID    329    385       PROBABLE.
FT   DISULFID    426    479       PROBABLE.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL)
FT   CONFLICT    215    215       Q -> R (IN REF. 2).
FT   CONFLICT    239    239       G -> R (IN REF. 2).
FT   CONFLICT    490    490       L -> F (IN REF. 2).
FT   CONFLICT    599    600       QG -> R (IN REF. 3).
FT   CONFLICT    720    721       MISSING (IN REF. 3).
FT   CONFLICT    811    811       G -> A (IN REF. 3).
SQ   SEQUENCE   848 AA;  93460 MW;  68D2F0C0E6C1C2AD CRC64;
```

```
  Query Match             12.4%;  Score 166;  DB 1;  Length 848;
  Best Local Similarity   28.0%;  Pred. No. 5.8e-06;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      49 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 94
           | :|  |                : :|  : ||: |    :| : ::|  ::  ||: : ||
Db     324 QVTLICEASGDPIPSIIWRTSTRNISSEEKTL----DGHMVVRSHARV--SSLILKSIQY 377

Qy      95 NDNGISFIC----RLGKDQSVSVSVVLNVIFPPLLSGNDFQIVEEGSNVKLVCNVKANPQ 150
           | |  : |      :|:|   | |: | | : | | |        ||: | : | | | |
Db     378 TDAG-EYICTASNTIGQD---SQSMYLEVQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 433

Qy     151 AQMMWYKNSSLPDLEK-SRHQIQQTISFSQISTLKVTEPDNGTYSCIAKSSLKTESLDFB 209
           | : |:::  ||     |  :| |  :  | :|   : | | |:| | : :  |||:|
Db     434 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 493

Qy     210 LIVKDKTVGVP----IEPIIAACVVIF 232
           |:  | |   |    :|| :   | |
Db     494 LVQAD-TPSSPSIDQVEPYSSTAQVQF 519
```

```
RESULT   7
NCAM_CHICK
ID   NCAM_CHICK     STANDARD;      PRT;  1091 AA.
AC   P13590; Q90919; Q90918;
DT   01-JAN-1990 (Rel. 13, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, large isoform precursor (N-CAM 180)
DE   (Contains: N-CAM 140).
OS   Gallus gallus (Chicken).
OC   Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
OC   Archosauria, Aves, Neognathae, Galliformes, Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-175 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=87206190; PubMed=3576199;
RA   Cunningham B.A., Hemperly J.J., Murray B.A., Prediger E.A.,
RA   Brackenbury R., Edelman G.M.;
RT   "Neural cell adhesion molecule: structure, immunoglobulin-like
RT   domains, cell surface modulation, and alternative RNA splicing.";
RL   Science 236:799-806(1987).
RN   [2]
RP   SEQUENCE OF 128-1091 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=86206089; PubMed=3458261;
RA   Hemperly J.J., Murray B.A., Edelman G.M., Cunningham B.A.;
RT   "Sequence of a cDNA clone encoding the polysialic acid rich and
RT   cytoplasmic domains of the neural cell adhesion molecule N-CAM.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:3037-3041(1986).
RN   [3]
RP   SEQUENCE FROM N.A. (726 AA FORM).
RX   MEDLINE=87092340; PubMed=3467341;
RA   Hemperly J.J., Edelman G.M., Cunningham B.A.;
RT   "cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a
RT   membrane-spanning region consistent with evidence for membrane
RT   attachment via a phosphatidylinositol intermediate.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:9822-9826(1986).
RN   [4]
RP   SEQUENCE OF 810-1069 FROM N.A.
RX   MEDLINE=87033934; PubMed=3771645;
RA   Murray B.A., Owens G.C., Prediger E.A., Crossin K.L.,
RA   Cunningham B.A., Edelman G.M.;
RT   "Cell surface modulation of the neural cell adhesion molecule
RT   resulting from alternative mRNA splicing in a tissue-specific
RT   developmental sequence.";
RL   J. Cell Biol. 103:1431-1439(1986).
RN   [5]
RP   SEQUENCE OF 1-17 FROM N.A.
RX   MEDLINE=93122797; PubMed=1478668;
RA   Colwell G., Li B., Forrest D., Brackenbury R.;
RT   "Conserved regulatory elements in the promoter region of the N-CAM
```

```
RT   gene.";
RL   Genomics 14:875-882(1992)
RN   [6]
RP   SEQUENCE OF 1-17 FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=Erythrocyte;
RA   Sasner M., Covault J.;
RL   Submitted (AUG-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15861; AAB59958.1; -.
DR   EMBL; M15860; AAB59958.1; JOINED.
DR   EMBL; M15922; AAB59958.1; JOINED.
DR   EMBL; M15923; AAB59958.1; JOINED.
DR   EMBL; M15924; AAB59958.1; JOINED.
DR   EMBL; M21178; AAB59958.1; JOINED.
DR   EMBL; M21179; AAB59958.1; JOINED.
DR   EMBL; M21180; AAB59958.1; JOINED.
DR   EMBL; M15929; AAB59958.1; JOINED.
DR   EMBL; M15930; AAB59958.1; JOINED.
DR   EMBL; M15931; AAB59958.1; JOINED.
DR   EMBL; M15932; AAB59958.1; JOINED.
DR   EMBL; M15933; AAB59958.1; JOINED.
DR   EMBL; M15934; AAB59958.1; JOINED.
DR   EMBL; L29437; AAB59958.1; JOINED.
DR   EMBL; M15935; AAB59958.1; JOINED.
DR   EMBL; M15937; AAB59958.1; JOINED.
DR   EMBL; M15938; AAB59958.1; JOINED.
DR   EMBL; M15939; AAB59958.1; JOINED.
DR   EMBL; M15861; AAB59959.1; -.
DR   EMBL; M15860; AAB59959.1; JOINED.
DR   EMBL; M15922; AAB59959.1; JOINED.
DR   EMBL; M15923; AAB59959.1; JOINED.
DR   EMBL; M15924; AAB59959.1; JOINED.
DR   EMBL; M21178; AAB59959.1; JOINED.
DR   EMBL; M21179; AAB59959.1; JOINED.
DR   EMBL; M21180; AAB59959.1; JOINED.
DR   EMBL; M15929; AAB59959.1; JOINED.
DR   EMBL; M15930; AAB59959.1; JOINED.
DR   EMBL; M15931; AAB59959.1; JOINED.
DR   EMBL; M15932; AAB59959.1; JOINED.
DR   EMBL; M15933; AAB59959.1; JOINED.
DR   EMBL; M15934; AAB59959.1; JOINED.
DR   EMBL; L29437; AAB59959.1; JOINED.
DR   EMBL; M15935; AAB59959.1; JOINED.
DR   EMBL; M15937; AAB59959.1; JOINED.
DR   EMBL; M15939; AAB59959.1; JOINED.
DR   EMBL; M15936; AAB59957.1; -.
DR   EMBL; M15860; AAB59957.1; JOINED.
DR   EMBL; M15922; AAB59957.1; JOINED.
DR   EMBL; M15923; AAB59957.1; JOINED.
DR   EMBL; M15924; AAB59957.1; JOINED.
DR   EMBL; M21178; AAB59957.1; JOINED.
DR   EMBL; M21179; AAB59957.1; JOINED.
DR   EMBL; M21180; AAB59957.1; JOINED.
DR   EMBL; M15929; AAB59957.1; JOINED.
DR   EMBL; M15930; AAB59957.1; JOINED.
DR   EMBL; M15931; AAB59957.1; JOINED.
DR   EMBL; M15932; AAB59957.1; JOINED.
DR   EMBL; M15934; AAB59957.1; JOINED.
DR   EMBL; L29437; AAB59957.1; JOINED.
DR   EMBL; M15935; AAB59957.1; JOINED.
DR   EMBL; X04479; CAB51638.1; -.
DR   EMBL; X70342; CAA49807.1; -.
DR   EMBL; Z12128; CAA78113.1; -.
DR   PIR; A43613; IJCHNL.
DR   HSSP; Q13740; 1KJC.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 4.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20   1091       NEURAL CELL ADHESION MOLECULE, LARGE
FT                                ISOFORM.
FT   DOMAIN       20    711       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    712    729       POTENTIAL.
FT   DOMAIN      730   1091       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    294       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      322    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      419    486       IG LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    595       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      624    692       FIBRONECTIN TYPE III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       BY SIMILARITY.
FT   DISULFID    139    189       BY SIMILARITY.
FT   DISULFID    235    287       BY SIMILARITY.
FT   DISULFID    329    385       BY SIMILARITY.
FT   DISULFID    426    479       BY SIMILARITY.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC     ) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    702    726       STSPTSGLGTAAIVGILIVIFVLLL -> TLGSPSTSSSFV
FT                                SLLLSAVILLLLC (IN 726 AA ISOFORM).
FT   VARSPLIC    727   1091       MISSING (IN 726 AA ISOFORM).
FT   VARSPLIC    771    809       MISSING (IN 1052 AA ISOFORM).
FT   VARSPLIC    810   1070       MISSING (IN ISOFORM N-CAM 140).
SQ   SEQUENCE   1091 AA;  117415 MW;  B151367002DF88E0 CRC64;


  Query Match          12.0%;  Score 160.5;  DB 1;  Length 1091;
  Best Local Similarity  26.9%;  Pred. No. 2.2e-05;
  Matches   60;  Conservative   37;  Mismatches   87;  Indels   39;  Gaps   10;

Qy      33 VNGKTENYILDTTPGSQASLIC-------------AVQNHTKEEELLWYKEEGRVDLKS 78
           |  ||   : |     | :| |               : :| : ||: |    :||: ::|
Db     313 VENKTAMELED-----QITLTCEASGDPIPSITWRTSTRNISNEEKTL----DGRIVVRS 363

Qy      79 GNKINSSSVCVSSISENDNGTSFIC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVE 134
             ::  ||: :  |   | |  : |     :|:|   | :: | | : | | |
Db     364 HARV--SSLTLKEIQYTDAG-EYVCTASNTIGQD---SQAMYLEVQYAPKLQGPVAVYTW 417

Qy     135 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTY 193
           ||: | : | | | | | : |::: ||     | :| |  :  | :|   : | | |
Db     418 EGNQVNITCEVFAYPSAVISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNY 477

Qy     194 SCIAKSSLKTESLDFHLIVKDKTVGVP----IEPIIAACVVIF 232
           :| | : :  || :| |:  | |   |    :|| :   | |
Db     478 NCTAVNRIGQESSEFILVQAD-TPSSPSIDRVEPYSSTARVEF 519
```

```
RESULT   8
NCA2_MOUSE
ID   NCA2_MOUSE     STANDARD;      PRT;   725 AA.
AC   P13594; Q61950;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, phosphatidylinositol-linked isoform
DE   precursor (N-CAM 120) (NCAM-120).
GN   NCAM1 OR NCAM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=87246524; PubMed=3595563;
RA   Barthels D., Santoni M.-J., Wille W., Ruppert C., Paix J.-C.,
RA   Hirsch M.-R., Fontecilla-Camps J.-C., Goridis C.;
RT   "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT   a Mr 79,000 polypeptide without a membrane-spanning region.",
RL   EMBO J. 6:907-914(1987).
RN   [2]
RP   SEQUENCE OF 20-700 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=89251561; PubMed=2721486;
RA   Santoni M.-J., Barthels D., Vopper G., Boned A., Goridis C., Wille W.;
RT   "Differential exon usage involving an unusual splicing mechanism
RT   generates at least eight types of NCAM cDNA in mouse brain.";
RL   EMBO J. 8:385-392(1989).
RN   [3]
RP   SEQUENCE OF 642-725 FROM N.A.
RX   MEDLINE=88283628; PubMed=3396534;
RA   Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.,
RT   "Differential splicing and alternative polyadenylation generates
RT   distinct NCAM transcripts and proteins in the mouse.";
RL   EMBO J. 7:625-632(1988).
RN   [4]
RP   SEQUENCE OF 20-36.
RX   MEDLINE=86140120; PubMed=3512556;
RA   Rougon G., Marshak D.R.;
RT   "Structural and immunological characterization of the amino-terminal
RT   domain of mammalian neural cell adhesion molecules.";
RL   J. Biol. Chem. 261:3396-3401(1986).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC    !  SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC    !- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ----------------------- ----------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; Y00051; CAA68263.1; -.
DR   EMBL; X15049; CAA33148.1; ALT_SEQ.
DR   EMBL; X07195; CAA30173.1; -.
DR   PIR; A29673; A29673.
DR   MGD; MGI:97281; Ncam.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
KW   Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor.
FT   SIGNAL       1     19
FT   CHAIN       20    725       NEURAL CELL ADHESION MOLECULE,
FT                                PHOSPHATIDYLINOSITOL-LINKED ISOFORM.
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    295       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      323    393       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      420    487       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      519    596       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      625    692       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       PROBABLE.
FT   DISULFID    139    189       PROBABLE.
FT   DISULFID    235    288       PROBABLE.
FT   DISULFID    330    386       PROBABLE.
FT   DISULFID    427    480       PROBABLE.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    348    348       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    424    424       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    450    450       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    479    479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    261    268       PRSRQSVS -> DPKHIFSD (IN REF. 2).
FT   CONFLICT    273    273       V -> L (IN REF. 2).
FT   CONFLICT    354    355       QD -> KT (IN REF. 2).
FT   CONFLICT    549    549       T -> K (IN REF. 2).
FT   CONFLICT    572    572       T -> R (IN REF. 2).
FT   CONFLICT    575    575       D -> V (IN REF. 2).
FT   CONFLICT    589    594       MQPSES -> SAATEF (IN REF. 2).
FT   CONFLICT    600    602       FEL -> REF (IN REF. 2).
FT   CONFLICT    657    657       H -> D (IN REF. 2 AND 3).
SQ   SEQUENCE   725 AA;  80296 MW;  C2AEB8B4461C6B2F CRC64;


  Query Match             11.9%;  Score 159;  DB 1;  Length 725;
  Best Local Similarity   27.5%;  Pred. No. 1.8e-05;
  Matches   57;  Conservative   35;  Mismatches   81;  Indels   34;  Gaps    9;

Qy      49 QASLIC--------------AVQNHTREEELLWYKEECRVDLKSCNKTNSSSVCVSSISE 94
           | :|  |               : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTILTCEASGDPIPSITWRTSTRNISSEEQDL----EGRMVVRSHARV--SSLTLKSIQY 378

Qy      95 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 150
             |  : |      :|:|   | |: |   : | | |       ||: | : | | | |
Db     379 EDAG-EYMCTASNTIGQD---SQSIHLEFQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     151 AQMMWYKNSSLIDFEK-SPHQLQQTSESPELNSILTVENPDNKRYSCIAKSSLKTESLDEH 209
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     210 LIVEDKIVSVE    IEPTAACVVIF 232
           |:  | |   |    :|| :   | |
Db     495 LVQADTPSSPSIDRVEPYSSTAQVQF 520


RESULT   9
NCA1_MOUSE
ID   NCA1_MOUSE     STANDARD;      PRT;  1115 AA.
AC   P13595; Q61949;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Neural cell adhesion molecule, large isoform precursor (N-CAM 180)
DE   (NCAM-180) [Contains: N-CAM 140 (NCAM-140)].
GN   NCAM1 OR NCAM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (N-CAM 180).
RC   STRAIN=C57BL/6;
RX   MEDLINE=87246524; PubMed=3595563;
RA   Barthels D., Santoni M.J., Wille W., Ruppert C., Chaix J.C.,
RA   Hirsch M.R., Fontecilla-Camps J.C., Goridis C.;
RT   "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT   a Mr 79,000 polypeptide without a membrane-spanning region.";
RL   EMBO J. 6:907-914(1987).
RN   [2]
RP   SEQUENCE OF 529-1115 FROM N.A. (N-CAM 140).
RC   STRAIN=C57BL/6;
RX   MEDLINE=88067687; PubMed=3684567;
RA   Santoni M.-J., Barthels D., Barbas J.A., Hirsch M.-R., Steinmetz M.,
RA   Goridis C., Wille W.;
RT   "Analysis of cDNA clones that code for the transmembrane forms of the
RT   mouse neural cell adhesion molecule (NCAM) and are generated by
RT   alternative RNA splicing.";
RL   Nucleic Acids Res. 15:8621-8641(1987)
RN   [3]
RP   SEQUENCE OF 642-1115 FROM N.A. (N-CAM 180)
RX   MEDLINE=88283628; PubMed=3396534;
RA   Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.;
RT   "Differential splicing and alternative polyadenylation generates
RT   distinct NCAM transcripts and proteins in the mouse.";
RL   EMBO J. 7:625-632(1988).
RN   [4]
RP   SEQUENCE OF 804-1081 FROM N.A. (N-CAM 180)
RC   STRAIN=C57BL/6J; TISSUE=Brain;
RX   MEDLINE=88247737; PubMed=2454455;
RA   Barthels D., Vopper G., Wille W.;
RT   "NCAM-180, the large isoform of the neural cell adhesion molecule of
RT   the mouse, is encoded by an alternatively spliced transcript.";
RL   Nucleic Acids Res. 16:4217-4225(1988)
RN   [5]
RP   SEQUENCE OF 702-1115 FROM N.A. (N-CAM 140).
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=89251563; PubMed=2721486;
RA   Santoni M.J., Barthels D., Vopper G., Boned A., Goridis C., Wille M.;
RT   "Differential exon usage involving an unusual splicing mechanism
RT   generates at least eight types of NCAM cDNA in mouse brain.";
RL   EMBO J. 8:385-392(1989).
RN   [6]
RP   SEQUENCE OF 20-36.
RX   MEDLINE=86140120; PubMed=3512556;
RA   Rougon G., Marshak D.R.;
RT   "Structural and immunological characterization of the amino-terminal
RT   domain of mammalian neural cell adhesion molecules.";
RL   J. Biol. Chem. 261:3396-3401(1986)
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC       NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC       NEURITES, ETC.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC       N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07200; CAA30177.1; -.
DR   EMBL; Y00051; -; NOT_ANNOTATED_CDS.
DR   EMBL; X06328; CAA29641.1; -.
DR   EMBL; X07195; CAA30173.1; -.
DR   EMBL; X07244; CAA30230.1; -.
DR   EMBL; X15051; CAA33150.1; -.
DR   EMBL; X15052; CAA33151.1; -.
DR   PIR; A29673; IJMSNL.
DR   MGD; MGI:97281; Ncam.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.
FT   SIGNAL        1     19
FT   CHAIN        20   1115       NEURAL CELL ADHESION MOLECULE, LARGE
FT                                 ISOFORM.
FT   DOMAIN       20    711       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    712    729       POTENTIAL.
FT   DOMAIN      730   1115       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       34    103       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    196       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      228    295       IG LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      323    393       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      420    487       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      519    596       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      625    692       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      152    156       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      161    165       HEPARIN-BINDING (POTENTIAL).
FT   DISULFID     41     96       PROBABLE.
FT   DISULFID    139    189       PROBABLE.
FT   DISULFID    235    288       PROBABLE.
FT   DISULFID    330    386       PROBABLE.
FT   DISULFID    427    480       PROBABLE.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    348    348       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    424    424       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    450    450       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    479    479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    810   1076       MISSING (IN ISOFORM N-CAM 140).
SQ   SEQUENCE   1115 AA;  119351 MW;  2C93DCD474CFBCAF CRC64;


  Query Match             11.9%;  Score 159;  DB 1;  Length 1115;
  Best Local Similarity   27.5%;  Pred. No. 3.1e-05;
  Matches   57;  Conservative   35;  Mismatches   81;  Indels   34;  Gaps    9;

Qy      49 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 94
           | :| |               : :| : ||: |    :| : ::|  ::  ||: : ||
Db     325 QVTLTCEASGDPIPSITWRTSTRNISSEEQDL----DGHMVVRSHARV--SSLTLKSIQY 378

Qy      95 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 150
             | |  : |     :|:|   | |: |   : | | |       ||: | : | | | |
Db     379 RDAG-EYMCTASNTIGQD---SQSIDLEFQYAPKLQGPVAVYTWEGNQVNITCEVFAYPS 434

Qy     151 AQMMWYKNESLLLDEK-SRMQTQQISESPQLSITKVEKPDNGTYSCIAKSSLNIESLDFH 209
           | : |:::  ||     |  :|  |  :  | :|   : | | |:| | : :  |||:|
Db     435 ATISWFRDGQLLPSSNYSNIKIYNTPSASYLEVTPDSENDFGNYNCTAVNRIGQESLEFI 494

Qy     210 LIVKDKTVGVP----IEPIIAACVVIF 232
           |:  | |   |    :||  :   | |
Db     495 LVQAD-TPSSPSIDRVEPYSSTAQVQF 520


RESULT  10
LAR_DROME
ID   LAR_DROME      STANDARD;      PRT;  2029 AA.
AC   P16621;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Protein-tyrosine phosphatase Lar precursor (EC 3.1.3.48) (Protein-
```

```
DE   tyrosine-phosphate phosphohydrolase) (DLAR).
GN   LAR.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90046860; PubMed=2554325;
RA   Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;
RT   "A family of receptor-linked protein tyrosine phosphatases in humans
RT   and Drosophila.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=96178473; PubMed=8598047;
RA   Krueger N.X., van Vactor D., Wan H.I., Gelbart W.M., Goodman C.S.,
RA   Saito H.;
RT   "The transmembrane tyrosine phosphatase DLAR controls motor axon
RT   guidance in Drosophila.";
RL   Cell 84:611-622(1996).
CC   -!- FUNCTION: IT IS POSSIBLE THAT DLAR IS A CELL ADHESION RECEPTOR.
CC       IT POSSESSES AN INTRINSIC PROTEIN TYROSINE PHOSPHATASE ACTIVITY
CC       (PTPASE). IT CONTROLS MOTOR AXON GUIDANCE.
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND
CC       PIONEER NEURONS IN THE EMBRYO.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 16 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27709; AAA28668.1; -.
DR   EMBL; U36857; AAC47002.1; -.
DR   EMBL; U36849; AAC47002.1; JOINED.
DR   EMBL; U36850; AAC47002.1; JOINED.
DR   EMBL; U36851; AAC47002.1; JOINED.
DR   EMBL; U36852; AAC47002.1; JOINED.
DR   EMBL; U36853; AAC47002.1; JOINED.
DR   EMBL; U36854; AAC47002.1; JOINED.
DR   EMBL; U36855; AAC47002.1; JOINED.
DR   EMBL; U36856; AAC47002.1; JOINED.
DR   PIR; A36182; TDFFLK.
DR   HSSP; P28827; 1RPM.
DR   FlyBase; FBgn0000464; Lar.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_prot_phphtase.
DR   Pfam; PF00041; fn3; 9.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00102; Y_phosphatase; 2.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00060; FN3; 9.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00194; PTPc; 2.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW   Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
KW   Cell adhesion; Immunoglobulin domain; Repeat.
FT   SIGNAL        1     32
FT   CHAIN        33   2029       PROTEIN-TYROSINE PHOSPHATASE LAR.
FT   DOMAIN       33   1377       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1378   1402       POTENTIAL.
FT   DOMAIN     1403   2029       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       50    118       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      154    216       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      249    308       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      320    417       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      418    512       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      513    607       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      608    706       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      707    809       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      810    906       FIBRONECTIN TYPE-III 6.
FT   DOMAIN      907   1007       FIBRONECTIN TYPE-III 7.
FT   DOMAIN     1008   1102       FIBRONECTIN TYPE-III 8.
FT   DOMAIN     1103   1207       FIBRONECTIN TYPE-III 9.
FT   DOMAIN     1492   1738       PROTEIN-TYROSINE PHOSPHATASE 1.
FT   DOMAIN     1781   2029       PROTEIN-TYROSINE PHOSPHATASE 2.
FT   ACT_SITE   1670   1670       BY SIMILARITY.
FT   ACT_SITE   1961   1961       BY SIMILARITY.
FT   DISULFID     57    111       POTENTIAL.
FT   DISULFID    161    209       POTENTIAL.
FT   DISULFID    256    301       POTENTIAL.
FT   CARBOHYD    176    176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    721    721       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    774    774       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    915    915       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    962    962       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1183   1183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1304   1304       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2029 AA;  229027 MW;  536A00794D3DC800 CRC64;
```

```
  Query Match           11.5%;  Score 153;  DB 1;  Length 2029;
  Best Local Similarity  25.9%;  Pred. No. 0.0002;
  Matches   51;  Conservative   29;  Mismatches   77;  Indels   40;  Gaps    9;

Qy      47 GSQASIICAVQNHTPPPFLIWKPPPGPVDEKSGNFINSSSVQVSSISPNDNGISF----T 102
           |  ||  || :       ::|         |:| |: | :    :: |   |||
Db      50 GGVASFYCAAEGDP-PPSIVW--------PRNGRKV-SGTQSRYIVIEQPGGISILRIEP 99

Qy     103 CRIGRD-------------QSVSVSVVLNV--------TPPPLLSGNDFQIVEEGSNVKL 141
           | |||              :||    | :          ||: |   : :| | | :
Db     100 VKAGRDDAPYECVAENGVGDAVSADATLIIYEGDKIPAGFPVIIQGPGTRVIEVGHTVLM 159

Qy     142 VCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSL 201
           |    ||   : | || : :|:   |: ::   : | | |    : | | | |:|:::|:
Db     160 ICKAIGNPTPNIYWIKNQTKVDMSNPPYSIK- -DGF-IQIENSPFEDQGKYECVAENSM 215

Qy     202 KTP-SLDFHLIVKDKTV 217
           || |   :| || : |
Db     216 GTEHSKATNLYVKVRRV 232


RESULT  11
PTPD_HUMAN
ID   PTPD_HUMAN     STANDARD;      PRT;  1912 AA.
AC   P23468;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Protein-tyrosine phosphatase delta precursor (EC 3.1.3.48) (R-PTP-
DE   delta).
GN   PTPRD.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.
RX   MEDLINE=95204468; PubMed=7896816;
RA   Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;
RT   "Molecular characterization of the human transmembrane protein-
RT   tyrosine phosphatase delta. Evidence for tissue-specific expression of
RT   alternative human transmembrane protein-tyrosine phosphatase delta
RT   isoforms.";
RL   J. Biol. Chem. 270:6722-6728(1995).
RN   [2]
RP   SEQUENCE OF 390-1912 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=91006018; PubMed=2170109;
RA   Krueger N.X., Streuli M., Saito H.;
RT   "Structural diversity and evolution of human receptor-like protein
RT   tyrosine phosphatases.";
RL   EMBO J. 9:3241-3252(1990).
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT
CC       TISSUES DUE TO ALTERNATIVE SPLICING.
CC   -!- PTM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN
CC       FROM THE TRANSMEMBRANE SEGMENT.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L38929; AAC41749.1; -.
DR   EMBL; X54133; CAA38068.1; -.
DR   PIR; S12052; S12052.
DR   HSSP; P18052; 1YFO.
DR   MIM; 601598; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_prot_phphtase.
DR   Pfam; PF00041; fn3; 8.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00102; Y_phosphatase; 2.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00060; FN3; 8
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00194; PTPc; 2.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW   Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane; Repeat;
KW   Immunoglobulin domain; Alternative splicing.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   1912       PROTEIN-TYROSINE PHOSPHATASE DELTA.
FT   DOMAIN      21   1265       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1266   1290       POTENTIAL.
FT   DOMAIN    1291   1912       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      23    115       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN     118    225       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     232    318       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN     320    414       FIBRONECTIN TYPE-III 1
FT   DOMAIN     417    513       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     516    606       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     609    708       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     711    822       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     825    916       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     918   1017       FIBRONECTIN TYPE-III 7.
FT   DOMAIN    1020   1137       FIBRONECTIN TYPE-III 8.
FT   DOMAIN    1375   1618       PROTEIN-TYROSINE PHOSPHATASE 1.
FT   DOMAIN    1619   1912       PROTEIN-TYROSINE PHOSPHATASE 2.
FT   ACT_SITE  1553   1553       BY SIMILARITY.
FT   ACT_SITE  1844   1844       BY SIMILARITY.
FT   SITE      1175   1178       CLEAVAGE (POTENTIAL).
FT   CARBOHYD   254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   724    724       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   832    832       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   181    189       MISSING (IN KIDNEY ISOFORM).
FT   VARSPLIC   226    229       MISSING (IN KIDNEY ISOFORM).
FT   VARSPLIC   775    785       MISSING (IN KIDNEY ISOFORM).
FT   VARSPLIC   609   1137       MISSING (IN FETAL BRAIN ISOFORM).
FT   MUTAGEN   1178   1178       R->A: 2.5-FOLD REDUCTION IN CLEAVAGE.
SQ   SEQUENCE  1912 AA;  214759 MW;  3AE8CBCD32182E26 CRC64;


  Query Match             11.2%;  Score 150;  DB 1;  Length 1912;
  Best Local Similarity   23.4%;  Pred. No. 0.00033;
  Matches   58;  Conservative   38;  Mismatches   90;  Indels   62;  Gaps   11;

Qy       8 IMQMGRFLLLVILFL-------PREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHT 60
           :: : | |||:: |         |   | : :   |           |  || ||
Db       1 MVHVAKLLLLLLIFFLKIDAEIPPREFKIPVDQIG---------VSGGVASFICQATGDP 51

Qy      61 PPEPLLWYPEEGRVDLKSGNKFNSSSVDVSSISFNDNGTSFTCPL-----GPDQS----- 110
           |  :::|         | | |:::     : | |:|     |:      ||::
Db      52 R-PKIVWN--------KKGKKVSNQRF---EVIEFDDGSGSVLRIQPLRTPRDEAIYECV 99

Qy     111 -------VSVSVVLNVT--------FPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMW 155
                 :|||  | |          || :  |   : ||      ::|    ||  :: |
Db     100 ASNNVGEISVSTRLTVLREDQIPRGFPTIDMGPQLKVVERTRTATMLCAASGNPDPEITW 159

Qy     156 YKNSSLLDLEKSPHQIQQ-TSFSF-------QLSITKVEPEDNGIYSCIANSSLKTIFSL 206
           :|:   :|   :  :|:|  |||         | | : |: | | | |:| :|  |  |
Db     160 FKDFLPVDTSNNNGRIKQLRSESIGGTPIRCALQIEQSEESDQGKYECVATNSAGTRYSA 219

Qy     207 DFHLIVKD 214
             :| |::
Db     220 PANLYVRE 227


RESULT  12
DCC_HUMAN
ID   DCC_HUMAN      STANDARD;      PRT;  1447 AA.
AC   P43146;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-MAR 2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).
GN   DCC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia, Eutheria, Primates, Catarrhini, Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95011532; PubMed=7926722;
RA   Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA   Vogelstein B.;
RT   "The DCC gene product in cellular differentiation and colorectal
RT   tumorigenesis.";
RL   Genes Dev. 8:1174-1183(1994).
RN   [2]
RP   SEQUENCE OF 1-750 FROM N.A.
RX   MEDLINE=90100559; PubMed=2294591;
RA   Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
```

```
RA   Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA   Vogelstein B.;
RT   "Identification of a chromosome 18q gene that is altered in
RT   colorectal cancers.";
RL   Science 247:49-56(1990).
RN   [3]
RP   SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).
RX   MEDLINE=91121517; PubMed=1991422;
RA   Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
RA   Oliner J.D., Kinzler K.W., Vogelstein B.;
RT   "Scrambled exons.";
RL   Cell 64:607-613(1991).
RN   [4]
RP   GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
RX   MEDLINE=94245241; PubMed=8188295;
RA   Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
RA   Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
RT   "The DCC gene: structural analysis and mutations in colorectal
RT   carcinomas.",
RL   Genomics 19:525-531(1994).
RN   [5]
RP   VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX   MEDLINE=94243823; PubMed=8187090;
RA   Miyake S., Nagai K., Yoshino K., Ota M., Endo M., Yuasa Y.;
RT   "Point mutations and allelic deletion of tumor suppressor gene DCC in
RT   human esophageal squamous cell carcinomas and their relation to
RT   metastasis.";
RL   Cancer Res. 54:3007-3010(1994).
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC       NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
CC   -!- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
CC       DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
CC       EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC       POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC       METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC    !  SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X76132; CAA53745.1; -.
DR   EMBL; M32292; AAA35751.1; -.
DR   EMBL; M32286; AAA52174.1; -.
DR   EMBL; M32288; AAA52175.1; ALT_SEQ.
DR   EMBL; M32290; AAA52176.1; -.
DR   EMBL; M63696; AAA52177.1; -.
DR   EMBL; M63700; AAA52178.1; -.
DR   EMBL; M63702; AAA52179.1; -.
DR   EMBL; M63718; AAA52180.1; -.
DR   EMBL; M63698; AAA52181.1; -.
DR   PIR; A54100; A54100.
DR   PIR; A40098; A40098.
DR   PIR; A38442; A38442.
DR   HSSP; P56276; 1TLK.
DR   MIM; 120470; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW   Repeat; Anti-oncogene, Disease mutation; Polymorphism.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1447       TUMOR SUPPPRESSOR PROTEIN DCC.
FT   DOMAIN       26   1097       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1098   1122       POTENTIAL.
FT   DOMAIN     1123   1447       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       54    124       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      154    219       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      254    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      345    407       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      426    522       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      525    618       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      619    716       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      722    816       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      840    940       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      941   1042       FIBRONECTIN TYPE-III 6.
FT   DISULFID     61    117       BY SIMILARITY.
FT   DISULFID    161    212       BY SIMILARITY.
FT   DISULFID    261    310       BY SIMILARITY.
FT   DISULFID    352    400       BY SIMILARITY.
FT   CARBOHYD     94     94       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC     ) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    628    628       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    702    702       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     168    168       M -> T (IN OESOPHAGEAL CARCINOMA).
FT                                /FTId=VAR_003909.
FT   VARIANT     201    201       R -> G.
FT                                /FTId=VAR_003910.
FT   VARIANT    1375   1375       P -> H (IN A COLORECTAL CARCINOMA).
FT                                /FTId=VAR_003911.
FT   CONFLICT    138    138       MISSING (IN REF. 3).
FT   CONFLICT    233    329       MISSING (IN REF. 3).
FT   CONFLICT    421    421       MISSING (IN REF. 3).
SQ   SEQUENCE   1447 AA;  158456 MW;  4A8612766ED0471F CRC64;
```

```
  Query Match             11.2%;  Score 149;  DB 1;  Length 1447;
  Best Local Similarity   25.8%;  Pred. No. 0.00028;
  Matches   51;  Conservative   37;  Mismatches   86;  Indels   24;  Gaps    9;

Qy      19 ILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKS 78
           : || |   |:|: : ||           | | | |  :     | | |  : |:|
Db     240 LYFLQR--PSNVVAIEGK-----------DAVLECCVSGYP-PPSFTWLRGEEVIQLRS 284

Qy      79 G--NKINSSSVCVSSISENDNGISFTCRLG-RDQSVSVSVVLNVTFPP-LLSGNDFQTVE 134
           : :  |:: :|:::::|:|: :|| :  :::::| |  | | ||  |:
Db     285 KKYSLEGGSNLLISNVIDDDSGM-YICVVIYKNENISASAELIVLVPPWFLNHPSNLYAY 343

Qy     135 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQISESFQLSTIKVEAPDNGIYS 194
           |  :::  | |   |   : | ||  :: :     ||   |    | |  | | | | |
Db     344 ESMDIEFECTVSGKPVPTVNWMKNGDVV-IPSDYFQIVGGS---NLRITGVVKSDEGFYQ 399

Qy     195 CIAKSSLKTESLDFHLIV 212
           |:|::        |||
Db     400 CVAENEAGNAQTSAQLIV 417


RESULT  13
LAMP_HUMAN
ID   LAMP_HUMAN     STANDARD;      PRT;   338 AA.
AC   Q13449;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Limbic system-associated membrane protein precursor (LSAMP).
GN   LSAMP OR LAMP.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96235133; PubMed=8666243;
RA   Pimenta A.F., Fischer I., Levitt P.;
RT   "cDNA cloning and structural analysis of the human limbic system
RT   associated membrane protein (LAMP).";
RL   Gene 170:189-195(1996).
CC   -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
CC       CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC       MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC       OF THE HYPPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON LIMBIC NEURONS AND FIBER TRACTS
CC       AS WELL AS IN SINGLE LAYERS OF THE SUPERIOR COLLICULUS, SPINAL
CC       CHORD AND CEREBELLUM
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U41901; AAC50569.1; -.
DR   MIM; 603241; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    315       LIMBIC SYSTEM-ASSOCIATED MEMBRANE
FT                                PROTEIN
FT   PROPEP      316    338       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       46    118       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      146    204       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      232    297       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID     53    111       POTENTIAL.
FT   DISULFID    153    197       POTENTIAL.
FT   DISULFID    239    290       POTENTIAL.
FT   CARBOHYD     40     40       N-LINKED (GLCNAC     ) (POTENTIAL)
FT   CARBOHYD     66     66       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    300    300       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       315    315       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   338 AA;  37308 MW;  03455F286DF5D92F CRC64;


  Query Match           11.1%;  Score 148.5;  DB 1;  Length 338;
  Best Local Similarity  25.4%;  Pred. No. 4.8e-05;
  Matches   60;  Conservative   34;  Mismatches   97;  Indels   45;  Gaps   12;

Qy      10 QMGRFLLLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR 69
           |:   || ::  ||  :    :  |  |:|  :     |  | | ::::  :  :: |
Db      11 QLPLVLLRLLCLLPTGLPVRSVDFNRGTDN--ITVRQGDTAILRCVLED--KNSKVAWLN 66

Qy      70 EEG------------RVDLKSGNKINSSSVCVSSISENDNGISFTCRLGRDQSVSVS-V 115
             |              ||:|:  : :   |: :  :   | | |:|| :       | |
Db      67 RSGIIFAGHDKWSLDPRVELEKRHSL-EYSLRIQKVDVYDEG-SYTCSVQTQHEPKTSQV 124

Qy     116 VLNVTFPPLLSG-NDFQIVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQT 174
           | |  || :|  :   || ||||| |||     |:  : |            || :  |
Db     125 YLIVQVPPKISNISSDVTVNEGSNVTLVCMANGRPEPVITW-----------RH-LTPT 171

Qy     175 SESFQ-----LSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGVPIEPII 225
             |:     | |  : :  :| | |  |:: :  | |    ||   | |   | |
Db     172 GREFEGEEEYLEILGITREQSGKYEC--KAANEVSSAD----VKQVKVTVNYPPTI 221


RESULT  14
PGBM_HUMAN
ID   PGBM_HUMAN     STANDARD;      PRT;  4393 AA.
AC   P98160; Q16287;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin, and Colon;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 1018-1472 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [4]
RP   SEQUENCE OF 892-1398 FROM N.A.
RC   TISSUE=Fibrosarcoma;
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1 >p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [5]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
CC   -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC       MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC       CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
```

```
CC       PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC       SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC       SUBSTRATE FOR CELLS.
CC   -!- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC       DIMERS OR STELLATE STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS
CC   -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC   -!- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; M64284; AAA52699.1; -.
DR   EMBL; S76446; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   HSSP; P01130; 1AJJ.
DR   Siena-2DPAGE; P98160; -.
DR   MIM; 142461; -.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001438; EGF_11.
DR   InterPro; IPR000006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 8.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00001; EGF_like; 8.
DR   SMART; SM00408; IGc2; 22.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   4393       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                               SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      197    236       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      283    321       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      323    361       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      366    405       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      405    506       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      523    532       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      533    732       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      733    765       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      766    815       LAMININ EGF-LIKE 2.
FT   DOMAIN      816    873       LAMININ EGF-LIKE 3.
FT   DOMAIN      881    925       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      926    935       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      936   1127       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1128   1160       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1161   1210       LAMININ EGF-LIKE 6.
FT   DOMAIN     1211   1267       LAMININ EGF-LIKE 7.
FT   DOMAIN     1277   1326       LAMININ EGF-LIKE 8.
FT   DOMAIN     1327   1336       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1337   1531       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1532   1564       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1565   1614       LAMININ EGF-LIKE 10.
FT   DOMAIN     1615   1672       LAMININ EGF-LIKE 11.
FT   DOMAIN     1679   1773       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     1774   1867       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN     1868   1957       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN     1958   2053       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN     2054   2153       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN     2154   2246       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN     2247   2342       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN     2343   2438       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN     2439   2535       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     2536   2631       IG-LIKE C2-TYPE DOMAIN 11.
FT   DOMAIN     2632   2728       IG-LIKE C2-TYPE DOMAIN 12.
FT   DOMAIN     2729   2828       IG-LIKE C2-TYPE DOMAIN 13.
FT   DOMAIN     2829   2926       IG-LIKE C2-TYPE DOMAIN 14.
FT   DOMAIN     2927   3023       IG-LIKE C2-TYPE DOMAIN 15.
FT   DOMAIN     3024   3114       IG-LIKE C2-TYPE DOMAIN 16.
FT   DOMAIN     3115   3213       IG-LIKE C2-TYPE DOMAIN 17.
FT   DOMAIN     3214   3300       IG-LIKE C2-TYPE DOMAIN 18.
FT   DOMAIN     3301   3401       IG-LIKE C2-TYPE DOMAIN 19.
FT   DOMAIN     3402   3490       IG-LIKE C2-TYPE DOMAIN 20.
FT   DOMAIN     3491   3576       IG-LIKE C2-TYPE DOMAIN 21.
FT   DOMAIN     3577   3664       IG-LIKE C2-TYPE DOMAIN 22.
FT   DOMAIN     3665   3845       LAMININ G-LIKE 1.
FT   DOMAIN     3846   3883       EGF-LIKE 1.
FT   DOMAIN     3886   3924       EGF-LIKE 2.
FT   DOMAIN     3930   4105       LAMININ G-LIKE 2.
FT   DOMAIN     4106   4143       EGF-LIKE 3.
FT   DOMAIN     4145   4178       EGF-LIKE 4.
FT   DOMAIN     4203   4391       LAMININ G-LIKE 3.
FT   SITE         65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE       4151   4153       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   SITE       4301   4303       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    206    225       BY SIMILARITY.
FT   DISULFID    219    234       BY SIMILARITY.
FT   DISULFID    285    297       BY SIMILARITY.
FT   DISULFID    292    310       BY SIMILARITY.
FT   DISULFID    304    319       BY SIMILARITY.
FT   DISULFID    325    337       BY SIMILARITY.
FT   DISULFID    332    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.

  Query Match          11.0%;  Score 147.5;  DB 1;  Length 4393;
  Best Local Similarity   26.2%;  Pred. No. 0.0015;
  Matches   55;  Conservative   38;  Mismatches   78;  Indels   39;  Gaps   12;

Qy      27 ISSVLEVNGETENYIEDITPGSQASIRCAVQNHTREFFFLWYRPFGRVDEKSGNKINSSS 86
           :||    || |   ||        |: :   ||    : ||: |   | | ::|  |
```

```
Db   2542 SSSASLANGHT----LDLN-----CLVASQAPHT----ITWYKRGG--SLPSRHQIVGSR 2586

Qy     87 VCVSSISENDNGISFTCRL-----GRDQSVSVSV-------VLNVTFPPLLSGNDFQTVE 134
          : :  ::  |:|  : | :      |: |: |::       |  |: ||:   :   ||
Db   2587 LRIPQVTPADSG-EYVCHVSNGAGSRETSLIVTIQGSGSSHVPRVS-PPIRIESSSPTVV 2644

Qy    135 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYS 194
          ||  : | | |   ||| : |||    |    ||||    :    | : ::   |:| |
Db   2645 EGQTLDLNCVVARQPQAIITWYKRGGSL---PSRHQ----THGSHLRLHQMSVADSGEYV 2697

Qy    195 CIAKSSLKTESLDFHLIVK-DKTVGVPIEP 223
          | | :::  ::|:  :::    : | |  |
Db   2698 CRANNNI--DALEASIVISVSPSAGSPSAP 2725


RESULT  15
DCC_MOUSE
ID   DCC_MOUSE      STANDARD;      PRT;  1447 AA.
AC   P70211;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor.
GN   DCC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX   NCBI TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Brain;
RX   MEDLINE=96112625; PubMed=8570174;
RA   Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F.;
RT   "Cloning of the mouse homologue of the deleted in colorectal cancer
RT   gene (mDCC) and its expression in the developing mouse embryo.";
RL   Oncogene 11:2243-2254(1995).
RN   [2]
RP   REVISIONS
RC   STRAIN=BALB/C; TISSUE=Brain;
RA   Cooper H.M.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM
CC       THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD
CC       FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC       DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC       BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC       EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC       AND REMAIN AT THIS LEVEL IN THE ADULT.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85788; CAA59786.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:94869; Dcc.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW   Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1447       TUMOR SUPPRESSOR PROTEIN DCC, LONG
FT                                ISOFORM.
FT   CHAIN        85   1447       TUMOR SUPPRESSOR PROTEIN DCC, SHORT
FT                                ISOFORM.
FT   INIT_MET     85     85       FOR SHORT ISOFORM.
FT   DOMAIN       26   1097       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1098   1122       POTENTIAL.
FT   DOMAIN     1123   1447       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       54    124       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      154    219       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      254    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      345    407       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      426    522       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      525    618       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      619    716       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      722    816       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      840    940       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      941   1042       FIBRONECTIN TYPE-III 6.
FT   DISULFID     61    117       BY SIMILARITY.
FT   DISULFID    161    212       BY SIMILARITY.
FT   DISULFID    261    310       BY SIMILARITY.
FT   DISULFID    352    400       BY SIMILARITY.
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     94     94       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    628    628       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    702    702       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    819    838       MISSING (IN EMBRYONIC ISOFORM).
SQ   SEQUENCE   1447 AA;  158298 MW;  0D1F1097C22D5B9F CRC64;


  Query Match             11.0%;  Score 147;  DB 1;  Length 1447;
  Best Local Similarity   25.8%;  Pred. No. 0.00041;
  Matches   51;  Conservative   37;  Mismatches   86;  Indels   24;  Gaps    9;

Qy      19 ILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKS 78
           : || |   |:|: : ||                | | | |  :     | | |  : |:|
Db    240 LYFLQR--PSNVIAIEGK------------DAVLECCVSGYP-PPSFTWLRGEEVIQLRS 284

Qy      79 G--NKINSSSVCVSSISENDNGISFTCKLG-RDQSVSVSVVLNVTFPP-LLSGNDFQTVE 134
             : :  |:: :|::::::|:| ::|| :  :::::| |  | |  ||  |:
Db    285 KKYSLLGGSNLLISNVTDDDSG-TYTCVVTYKNENISASAELTVLVPPWFLNHPSNLYAY 343

Qy    135 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYS 194
          |  :::  | |   |   : | ||  :: :     ||   |    | |  | | | | |
Db    344 ESMDIEFECAVSGKPVPTVNWMKNGDVV-IPSDYFQIVGGS---NLRILGVVKSDEGFYQ 399

Qy    195 CIAKSSLKTESLDFHLIV 212
          |:|::          |||
Db    400 CVAENEAGNAQSSAQLIV 417


Search completed: September 17, 2002, 13:24:17
Job time: 804 sec
```

GenCore version 4.5
Copyright (c) 1993 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 08:21:32 ; Search time 2925.1 Seconds
(without alignments)
5472.907 Million cell updates/sec

Title: US-09-863-823-1
Perfect score: 765
Sequence: 1 atgcaaatgggaagatttct.....acagtgaaacagctctatga 765

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1048326829[illegible] residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	504.6	66.0	1340	10	BC022683	BC022683 Mus muscu
c 2	286.6	37.5	70248	2	AC103795	AC103795 Homo sapi
c 3	286.6	37.5	171136	9	AC090685	AC090685 Homo sapi
c 4	286.6	37.5	172046	2	AC087644	AC087644 Homo sapi
c 5	284	37.1	172622	2	AC024065	AC024065 Homo sapi
6	277.4	36.3	338	6	AX245713	AX245713 Sequence
7	227.2	29.7	361	6	AX246249	AX246249 Sequence
8	225.4	29.5	328	6	AX246408	AX246408 Sequence
9	223.6	29.2	359	6	AX070657	AX070657 Sequence
10	214.6	28.1	201797	2	AL645479	AL645479 Mus muscu
c 11	214.6	28.1	229289	10	AL603842	AL603842 Mouse DNA
12	179.8	23.5	359	6	AX070653	AX070653 Sequence
13	109.6	14.3	70248	2	AC103795	AC103795 Homo sapi
c 14	43.8	5.7	217730	2	AC087440	AC087440 Homo sapi
c 15	42	5.5	125020	9	AF429315	AF429315 Homo sapi
16	42	5.5	217730	2	AC087440	AC087440 Homo sapi
c 17	40.2	5.3	393	11	G37560	G37560 SHGC-57994
18	39.6	5.2	2134	3	DMU51473	U51473 Drosophila
19	39.6	5.2	8500	2	AC023864	AC023864 Drosophil
20	39.6	5.2	68727	3	AC004516	AC004516 Drosophil
c 21	39.6	5.2	193262	3	AC007579	AC007579 Drosophil
c 22	39.6	5.2	262173	3	AE003808	AE003808 Drosophil
23	39.2	5.1	2799	9	HSNCAME	X16841 Human mRNA
24	39.2	5.1	2960	9	S71824	S71824 N-CAM-145 k
25	39.2	5.1	3309	9	AK057509	AK057509 Homo sapi
26	39.2	5.1	4092	6	AK062746	AK062746 Sequence
27	39.2	5.1	4092	9	HUMKOR2A	M97639 Human trans
28	39.2	5.1	8513	6	AX207284	AX207284 Sequence
29	39	5.1	172567	2	AC015493	AC015493 Homo sapi
c 30	39	5.1	203252	2	AC025421	AC025421 Homo sapi
31	38.8	5.1	157756	2	AP001075	AP001075 Homo sapi
c 32	38.8	5.1	195622	2	AP000807	AP000807 Homo sapi
33	38.6	5.0	2855	10	RNU34985	U34985 Rattus norv
34	38.6	5.0	2869	6	A72401	A72401 Sequence 2
35	38.6	5.0	2869	6	AR043363	AR043363 Sequence
36	38.6	5.0	2869	6	I60497	I60497 Sequence 2
c 37	38.6	5.0	65119	9	AL590062	AL590062 Human DNA
c 38	38.6	5.0	125698	9	HS692C8	AL034561 Human DNA
c 39	38.6	5.0	159287	9	AP003694	AP003694 Homo sapi
40	38.6	5.0	175218	2	AC027031	AC027031 Homo sapi
c 41	38.6	5.0	177731	2	AC018937	AC018937 Homo sapi
c 42	38.4	5.0	291	1	MC385	Z33288 M.capricolu
c 43	38.2	5.0	118368	9	AC079612	AC079612 Homo sapi
44	38	5.0	71023	2	AC004426	AC004426 Drosophil
c 45	38	5.0	106585	9	AL160392	AL160392 Human DNA

ALIGNMENTS

RESULT 1
BC022683
LOCUS BC022683 1340 bp mRNA linear EST 24-FEB-2002
DEFINITION Mus musculus, RIKEN cDNA 2010002A20 gene, clone IMAGE:4241194, mRNA, partial cds.
ACCESSION BC022683
VERSION BC022683.1 GI:18490500
KEYWORDS .
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1340)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) med@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 43 Row: o Column: 9.
FEATURES              Location/Qualifiers
     source           1. .1340
                      /organism="Mus musculus"
                      /db_xref="LocusID:66601"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:4241194"
                      /tissue_type="Kidney, normal, 5 month old male mouse."
                      /clone_lib="NCI_CGAP_Kid14"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     CDS              <1. .776
                      /codon_start=3
                      /product="RIKEN cDNA 2010002A20 gene"
                      /protein_id="AAH22683.1"
                      /db_xref="GI:18490501"
                      /translation="ITGPLQACQLLLVVLSLPGGRTSSVLTVNGPTENYIIEGQHGVQ
                      ASLECAVQNHIEDEELLWYRDGIVDLKNGNKINISSVCVSPINESDNGVKFFCKIQR
                      DQTVSVTVVLNVTFPPLLSGNGFQTVEENSDVSLVCNVKSNPQAQMMWYKNNSALVLE
                      KGRHQIHQTRESFQLSITKVKKSDNGTYSCIASSSLKMETMDFHLLVKDKVFVMPAEP
                      IIAACVVVVLTMAFALFSKRKRIMKLCGKANDPNSETAL"
BASE COUNT       418 a    261 c    288 g    373 t
ORIGIN


  Query Match             66.0%;  Score 504.6;  DB 10;  Length 1340;
  Best Local Similarity   79.5%;  Pred. No. 5.5e-138;
  Matches  594;  Conservative    2;  Mismatches  151;  Indels    0;  Gaps    0;

Qy      19 cttctcttagtaatttttatttctgccacgtgagatgacaagttctgttttaactgtgaat 78
           ||||||||||||  |||||| | ||||||  | ||  ||||| ||||||||||||||||||
Db      30 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGGAGAACAAGCTCTGTTTTAACTGTGAAT 89

Qy      79 gqtaaaactgagaactatatctgqatactacacctgqctcccaaycatctctgatatgt 138
           |||| ||||||||||||||| ||||| ||| || ||| ||||||||||||  ||||
Db      90 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 149

Qy     139 qctqttcaaaaccacaccagagaqqaaqaaactqctctqgtaccqagaggaggqqqaqagtg 198
           ||||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db     150 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAAGATGGAATAGTA 209

Qy     199 qatttqaaatctqaaaacaaaatcaattcaqctctqtctqtctcttccatcaqtqaa 258
           ||| |||||  ||:|||||:|||||   ||| ||||| ||||||||| ||||||  |||
Db     210 GATCTGAAAAATGCAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAA 269

Qy     259 aatqacaaccqqaatcaqctttacctqcaggctqqqqaqqgatcaqtccqtqtccgtttcq 318
           | |||||| ||| || ||||||||||| |||   | | |||||| | |||||||| |
Db     270 AGTCACAATGCAGTCCGCTTTACCTGCAAGCTACAGCCGGATCAGACGGTGTCCGTTACA 329

Qy     319 qtqqtqctqaatqttacttttcctcctctcctaaqtqqaaacqacttccaaacaqttqaq 378
           || |||||||| ||||| |||||||||||| |||||||| || | ||||||||| || |||
Db     330 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAG 389

Qy     379 qaaqqcqtqatqtqaaqttqqtttqcaatqtqaaaqcccccccccagqctcaaatqatq 438
           |||  |||  |||| |  |||||||||||| |||||  ||||||| || || |||||||||
Db     390 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 449

Qy     439 tggtacaaaaacaqtaqtctcctcqatttaqaqaaaaqccqtcaccaaatccaacacqaca 498
           ||||| |||||||  ||   | | | |||||||||| ||||||||||||||||| ||||||
Db     450 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACA 509

Qy     499 agtqaqtctttcaqctqtcaatcaccaaaqtcqaqaaqcctqacaacqqaacctacagt 558
           .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
           || |||||||||||| |||| |||||||||||| ||||  ||||||| || ||||||||
Db    510 AGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 569

Qy    559 tqtattqcaaaqtcatctctqaaaaqqqaqaqcttqqactttcaccqtqattqttaaqat 618
          |||||||| |  |||||||||||||| ||||| | ||||||| |||||| ||||||||||
Db    570 TGTATTGCCAGCTCATCTCTGAAAATCCAGACCATCCACTTCCACCTGCTTGTTAAAGAC 629

Qy    619 aaaactqtgggtgtaccaatagagcccattattgctqcatqtqttqtgatcttctqaca 678
          |||  | | |   | ||   |||||||| |||||||| ||||| ||| || | |||||
Db    630 AAAGTTTTCGTCATGCCGGCAGACCCATCATTGCTGCTTGTGTCGTGGTCGTCGTCTGACG 689

Qy    679 ttqtqctttqqactqattgctaqaaqaaaqaaaataatqaagctctgcatqaaqqataaa 738
          ||   |||| ||| || |||| || || | ||||||||||||||||   ||| | ||
Db    690 ATGGCTTTGCCCTGTTTTCTAGGAGGAAAAGAATAATGAAGCTCTGTGGGAAAAAGAAT 749

Qy    739 gaccctcacaqtqaaacagctctatqa 765
          |||||  |    ||||||||||||| |
Db    750 GACCCGAATTCAGAAACAGCTCTATAA 776


RESULT   2
AC103795/c
LOCUS       AC103795               70248 bp    DNA     linear   HTG 29-NOV-2001
DEFINITION  Homo sapiens chromosome 11 clone CTD-2529D2 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC103795
VERSION     AC103795.1  GI:17149680
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70248)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone CTD-2529D2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70248)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,P., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zaitoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21770
                Center clone name: 2529_D_2
```

```
--------------.
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      687: contig of 687 bp in length
*      688 787: gap of      100 bp
*      788     1501: contig of 714 bp in length
*     1502 1601: gap of      100 bp
*     1602     2345: contig of 744 bp in length
*     2346 2445: gap of      100 bp
*     2446     3151: contig of 706 bp in length
*     3152 3251: gap of      100 bp
*     3252     3929: contig of 678 bp in length
*     3930 4029: gap of      100 bp
*     4030     4719: contig of 690 bp in length
*     4720 4819: gap of      100 bp
*     4820     5538: contig of 719 bp in length
*     5539 5638: gap of      100 bp
*     5639     6356: contig of 718 bp in length
*     6357 6456: gap of      100 bp
*     6457     7156: contig of 700 bp in length
*     7157 7256: gap of      100 bp
*     7257     7888: contig of 632 bp in length
*     7889 7988: gap of      100 bp
*     7989     8743: contig of 755 bp in length
*     8744 8843: gap of      100 bp
*     8844     9533: contig of 690 bp in length
*     9534 9633: gap of      100 bp
*     9634    10371: contig of 738 bp in length
*    10372 10471: gap of      100 bp
*    10472    11203: contig of 732 bp in length
*    11204 11303: gap of      100 bp
*    11304    12016: contig of 713 bp in length
*    12017 12116: gap of      100 bp
*    12117    12823: contig of 707 bp in length
*    12824 12923: gap of      100 bp
*    12924    13628: contig of 705 bp in length
*    13629 13728: gap of      100 bp
*    13729    14441: contig of 713 bp in length
*    14442 14541: gap of      100 bp
*    14542    15272: contig of 731 bp in length
*    15273 15372: gap of      100 bp
*    15373    16106: contig of 734 bp in length
*    16107 16206: gap of      100 bp
*    16207    16935: contig of 729 bp in length
*    16936 17035: gap of      100 bp
*    17036    17762: contig of 727 bp in length
*    17763 17862: gap of      100 bp
*    17863    18558: contig of 696 bp in length
*    18559 18658: gap of      100 bp
*    18659    19351: contig of 693 bp in length
*    19352 19451: gap of      100 bp
*    19452    20181: contig of 730 bp in length
*    20182 20281: gap of      100 bp
*    20282    21023: contig of 742 bp in length
*    21024 21123: gap of      100 bp
*    21124    21827: contig of 704 bp in length
*    21828 21927: gap of      100 bp
*    21928    22624: contig of 697 bp in length
*    22625 22724: gap of      100 bp
*    22725    23401: contig of 677 bp in length
*    23402 23501: gap of      100 bp
*    23502    24228: contig of 727 bp in length
*    24229 24328: gap of      100 bp
*    24329    25055: contig of 727 bp in length
*    25056 25155: gap of      100 bp
*    25156    25898: contig of 743 bp in length
*    25899 25998: gap of      100 bp
*    25999    26710: contig of 712 bp in length
*    26711 26810: gap of      100 bp
*    26811    27551: contig of 741 bp in length
*    27552 27651: gap of      100 bp
*    27652    28394: contig of 743 bp in length
*    28395 28494: gap of      100 bp
*    28495    29230: contig of 736 bp in length
*    29231 29330: gap of      100 bp
*    29331    30037: contig of 707 bp in length
*    30038 30137: gap of      100 bp
*    30138    30854: contig of 717 bp in length
*    30855 30954: gap of      100 bp
*    30955    31680: contig of 726 bp in length
*    31681 31780: gap of      100 bp
*    31781    32520: contig of 740 bp in length
*    32521 32620: gap of      100 bp
*    32621    33354: contig of 734 bp in length
*    33355 33454: gap of      100 bp
*    33455    34191: contig of 737 bp in length
*    34192 34291: gap of      100 bp
*    34292    35034: contig of 743 bp in length
*    35035 35134: gap of      100 bp
*    35135    35880: contig of 746 bp in length
*    35881 35980: gap of      100 bp
*    35981    36663: contig of 683 bp in length
*    36664 36763: gap of      100 bp
*    36764    37477: contig of 714 bp in length
*    37478 37577: gap of      100 bp
*    37578    38321: contig of 744 bp in length
*    38322 38421: gap of      100 bp
*    38422    39110: contig of 689 bp in length
*    39111 39210: gap of      100 bp
*    39211    39892: contig of 682 bp in length
*    39893 39992: gap of      100 bp
*    39993    40678: contig of 686 bp in length
*    40679 40778: gap of      100 bp
*    40779    41479: contig of 701 bp in length
*    41480 41579: gap of      100 bp
*    41580    42310: contig of 731 bp in length
*    42311 42410: gap of      100 bp
*    42411    43144: contig of 734 bp in length
*    43145 43244: gap of      100 bp
*    43245    43985: contig of 741 bp in length
*    43986 44085: gap of      100 bp
*    44086    44777: contig of 692 bp in length
*    44778 44877: gap of      100 bp
*    44878    45598: contig of 721 bp in length
*    45599 45698: gap of      100 bp
*    45699    46433: contig of 735 bp in length
*    46434 46533: gap of      100 bp
*    46534    47267: contig of 734 bp in length
*    47268 47367: gap of      100 bp
*    47368    48058: contig of 691 bp in length
*    48059 48158: gap of      100 bp
*    48159    48880: contig of 722 bp in length
*    48881 48980: gap of      100 bp
*    48981    49717: contig of 737 bp in length
*    49718 49817: gap of      100 bp
*    49818    50560: contig of 743 bp in length
*    50561 50660: gap of      100 bp
*    50661    51400: contig of 740 bp in length
*    51401 51500: gap of      100 bp
*    51501    52227: contig of 727 bp in length
*    52228 52327: gap of      100 bp
*    52328    53066: contig of 739 bp in length
*    53067 53166: gap of      100 bp
*    53167    53853: contig of 687 bp in length
*    53854 53953: gap of      100 bp
*    53954    54675: contig of 722 bp in length
*    54676 54775: gap of      100 bp
```

```
            *     54776    55508: contig of 733 bp in length
            *     55509 55608: gap of      100 bp
            *     55609    56347: contig of 739 bp in length

  Query Match             37.5%;  Score 286.6;  DB 2;  Length 70248;
  Best Local Similarity   97.0%;  Prod  No  1.8e-73;
  Matches  292;  Conservative    9,  Mismatches    9,  Indels    0,  Caps    0,

Qy      326 tgaatgttactttttctctctctctaagtggaaaacgacttccaaaacagttgaggaaggca 385
            ||| | || |  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49431 TGAGTCTTGCAGTTCCTCCTCTCTAAGTGGAAACGACTTCCAAACAGTTGAGGAAGGCA 49372

Qy      386 gtaatgtgaaagttggtttgcaatgtgaaaagccaaccccccaggctcaaatgatgtggtaca 445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49371 GTAATGTGAAGTTGGTTTGCAATGTGAAAGCCAAGCCCCAGGCTCAAATGATGTGGTACA 49312

Qy      446 aaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagt 505
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49311 AAAACAGTAGTCTCCTCGATTTAGAGAAAAGCCGTCACCAAATCCAACAGACAAGTGAGT 49252

Qy      506 cttttcagctgtcaatcaccaaagtcagagaagcctgacaacggaacctacagttgtattg 565
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49251 CTTTTCAGCTGTCAATCACCAAAGTCGAGAAGCCTGACAACGGAACCTACAGTTGTATTG 49192

Qy      566 caaagtcatctctgaaaacggagagctttggacttttcacctgattgttaaaggtaatgctc 625
            ||||||||||||||||||||||||||||||||||||||||||||||||| |||  ||||||||
Db    49191 CAAAGTCATCTCTGAAAACGGAGAGCTTTGGACTTTCACCTGATTGTTAAAGGTAATGCTC 49134

Qy      626 t 626
            |
Db    49133 T 49133


RESULT   3
AC090685/c
LOCUS       AC090685              171136 bp    DNA     linear   PRI 27-SEP-2001
DEFINITION  Homo sapiens chromosome 17, clone RP11-252O24, complete sequence.
ACCESSION   AC090685
VERSION     AC090685.7  GI:15789269
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171136)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11 252O24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171136)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boquslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Gindo,S., Goyette,M., Graham,L., Grand Pierre,N.,
            Haqos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenqa,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Souqnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171136)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boquslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Gindo,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Haqos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenqa,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27 SEP 2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 27, 2001 this sequence version replaced gi:15421875.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10758
                Center clone name: 252_O_24
            --------------.
FEATURES             Location/Qualifiers
     source          1. .171136
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11 252O24"
                     /clone_lib="RPCI 11 Human Male BAC"
     repeat_region   496. .810
                     /rpt_family="AluSx"
     repeat_region   complement(903. .1220)
                     /rpt_family="AluSx"
     repeat_region   1958. .2259
                     /rpt_family="AluSx"
     repeat_region   3521. .3541
                     /rpt_family="AT_rich"
     repeat_region   complement(4906. .5206)
                     /rpt_family="AluSx"
     repeat_region   6440. .6732
                     /rpt_family="AluJb"
     repeat_region   6737. .6776
                     /rpt_family="MER4E"
     repeat_region   6794. .6971
                     /rpt_family="AluJb"
     repeat_region   7119. .7242
                     /rpt_family="AluSq"
     repeat_region   7243. .7273
                     /rpt_family="(CAAAA)n"
     repeat_region   7274. .7440
                     /rpt_family="AluSq"
     repeat_region   complement(7666. .7865)
                     /rpt_family="MIR"
     repeat_region   complement(7983. .8112)
                     /rpt_family="MIR"
```

```
repeat_region   complement(8499. .8785)
                /rpt_family="L1ME3"
repeat_region   9605. .9645
                /rpt_family="AT_rich"
repeat_region   10172. .10398
                /rpt_family="L2"
repeat_region   complement(10430. .10704)
                /rpt_family="AluSx"
repeat_region   11396. .11560
                /rpt_family="AluJb"
repeat_region   complement(12217. .12517)
                /rpt_family="AluJo"
repeat_region   12807. .12849
                /rpt_family="(T)n"
repeat_region   13056. .13161
                /rpt_family="MER5A"
repeat_region   13340. .13409
                /rpt_family="L2"
repeat_region   complement(13413. .13530)
                /rpt_family="FRAM"
repeat_region   complement(13613. .13679)
                /rpt_family="MER5B"
repeat_region   complement(13811. .13933)
                /rpt_family="FLAM_A"
repeat_region   complement(13960. .14253)
                /rpt_family="AluSx"
repeat_region   14620. .14901
                /rpt_family="AluJo"
repeat_region   complement(15259. .15559)
                /rpt_family="AluSc"
repeat_region   complement(15645. .15871)
                /rpt_family="MIR"
repeat_region   complement(15948. .16005)
                /rpt_family="L2"
repeat_region   complement(16005. .16129)
                /rpt_family="MIR"
repeat_region   16580. .16864
                /rpt_family="L2"
repeat_region   18380. .18407
                /rpt_family="AT_rich"
repeat_region   complement(18780. .19066)
                /rpt_family="AluJo"
repeat_region   19130. .19163
                /rpt_family="U2"
repeat_region   complement(19812. .19917)
                /rpt_family="U6"
repeat_region   complement(20095 .20376)
                /rpt_family="AluJb"
repeat_region   20423. .20727
                /rpt_family="AluSx"
repeat_region   20848. .20872
                /rpt_family="AT_rich"
repeat_region   23695. .23721
                /rpt_family="(CGG)n"
repeat_region   complement(25012. .25643)
                /rpt_family="L2"
repeat_region   complement(25677. .25939)
                /rpt_family="L2"
repeat_region   25998. .26149
                /rpt_family="MER91A"
repeat_region   complement(26150. .26443)
                /rpt_family="AluSq"
repeat_region   26444. .26493
                /rpt_family="MER91A"
repeat_region   complement(26502. .26750)
                /rpt_family="L2"
repeat_region   27314. .27582
                /rpt_family="AluJo"
repeat_region   complement(27786. .28088)
                /rpt_family="AluSq"
repeat_region   28275. .28362
                /rpt_family="Charlie8"
repeat_region   28363. .28641
                /rpt_family="AluJb"
repeat_region   28642. .28715
                /rpt_family="Charlie8"
repeat_region   28729. .28827
                /rpt_family="AluJb"
repeat_region   28828. .29867
                /rpt_family="LTR28"
repeat_region   29868. .30069
                /rpt_family="AluSq"
repeat_region   complement(30661. .30743)
                /rpt_family="L2"
repeat_region   complement(31351. .31487)
                /rpt_family="L2"
repeat_region   31524. .31656
                /rpt_family="AluSq/x"
repeat_region   complement(31657. .31951)
```

```
  Query Match             37.5%;  Score 286.6;  DB 9;  Length 171136;
  Best Local Similarity   97.0%;  Pred. No. 1.9e-73;
  Matches  292;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     326 tgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggca 385
           ||| | || |  ||||||||||||||||||||||||||||||||||||||||||||||||
Db   56860 TGAGTCTTGCAGTTCCTCCTCTCCTAAGTGGAAACGACTTCCAAACAGTTGAGGAAGGCA 56801

Qy     386 gtaatgtgaagttggtttgcaatgtgaaagccaacccccaggctcaaatgatgtggtaca 445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56800 GTAATGTGAAGTTGGTTTGCAATGTGAAAGCCAACCCCCAGGCTCAAATGATGTGGTACA 56741

Qy     446 aaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagt 505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56740 AAAACAGTAGTCTCCTCGATTTAGAGAAAAGCCGTCACCAAATCCAACAGACAAGTGAGT 56681

Qy     506 cttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattg 565
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56680 CTTTTCAGCTGTCAATCACCAAAGTCGAGAAGCCTGACAACGGAACCTACAGTTGTATTG 56621

Qy     566 caaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagataaaactg 625
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||  ||
Db   56620 CAAAGTCATCTCTGAAAACGGAGAGCTTGGACTTTCACCTGATTGTTAAAGGTAATGCTC 56561

Qy     626 t 626
           |
Db   56560 T 56560
```

```
RESULT   4
AC087644/c
LOCUS       AC087644               172046 bp    DNA     linear   HTG 06-JAN-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-403E9 map 17, *** SEQUENCING
            IN PROGRESS ***, 8 unordered pieces.
ACCESSION   AC087644
VERSION     AC087644.3  GI:18072182
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota, Metazoa, Chordata, Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172046)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-403E9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172046)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
```

McPherson,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Ruy,A., Santos,R., Schauer,S., Schupback,E., Seaman,S., Severy,F., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Triqillio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,S., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jan 6, 2002 this sequence version replaced gi:13357327.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L11984
Center clone name: 403_E_9

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 47343: contig of 47343 bp in length
* 47344 47443: gap of 100 bp
* 47444 51984: contig of 4541 bp in length
* 51985 52084: gap of 100 bp
* 52085 56314: contig of 4230 bp in length
* 56315 56414: gap of 100 bp
* 56415 61504: contig of 5090 bp in length
* 61505 61604: gap of 100 bp
* 61605 64448: contig of 2844 bp in length
* 64449 64548: gap of 100 bp
* 64549 66982: contig of 2434 bp in length
* 66983 67082: gap of 100 bp
* 67083 69497: contig of 2415 bp in length
* 69498 69597: gap of 100 bp
* 69598 172046: contig of 102449 bp in length.

FEATURES Location/Qualifiers
source 1..172046
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-403E9"
/clone_lib="RPCI-11 Human Male BAC"

BASE COUNT 50159 a 36414 c 35904 g 48710 t 859 others

ORIGIN

Query Match 47.5%; Score 286.6; DB 2; Length 172046;
Best Local Similarity 97.0%; Pred. No. 1.9e-73;
Matches 292; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy     326 tgaatatttacttttcctctctctaaagtggaaacgacttccaaacagttgaggaaggca 385
           ||| | || |  ||||||||||||||||||||||||||||||||||||||||||||||||
Db 154978 TGAGTCTTGCAGTTGCTCCTCTCCTAAGTGGAAACGACTTCCAAACAGTTGAGGAAGGCA 154919

Qy     386 gtaatatgtaaagttgattgcaatatgtgaaagccaaccccccaggctcaaatgatgtggtaca 445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154918 GTAATGTGAAGTTGGTTTGCAATGTGAAAGCCAACCCCCAGGCTCAAATGATGTGGTACA 154859

Qy     446 aaaacagtagtctcctcgatttagagaaaagcccgtcaccaaatccaacagacaagtgagt 505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154858 AAAACAGTAGTCTCCTCGATTTAGAGAAAAGCCCGTCACCAAATCCAACAGACAAGTGAGT 154799

Qy     506 ctttttagctgtcaatcacaaaagtcgagaagcctgaaaacggaacctacagttgtattg 565
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154798 CTTTTCAGCTGTCAATCACCAAAGTCGAGAAGCCTGAAAACGGAACCTACAGTTGTATTG 154739

Qy     566 caaagtcatctctgaaaacggagagcttggactttcacctgattgttaaaggtaatgctc 625
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||
Db 154738 CAAAGTCATCTCTGAAAACGGAGAGCTTGGACTTTCACCTGATTGTTAAAGGTAATGCTC 154679

Qy     626 t 626
           |
Db 154678 T 154678
```

RESULT 5
AC024065/c
LOCUS AC024065 172622 bp DNA linear HTG 17-AUG-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-403E9, WORKING DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION AC024065
VERSION AC024065.5 GI:9838183
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172622)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172622)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:9502467.

---------- Genome Center ----------
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
---------- Project Information ----------
Center project name: H_NH0403E09
---------- Summary Statistics ----------
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 167048 bases at least Q40
Consensus quality: 169424 bases at least Q30
Consensus quality: 170452 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 173251; sum-of-contigs
Quality coverage: 5.19 in Q20 bases; agarose-fp
Quality coverage: 5.88 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6930: contig of 6930 bp in length
* 6931 7030: gap of unknown length
* 7031 24747: contig of 17717 bp in length

```
            *    24748    24847: gap of unknown length
            *    24848    35273: contig of 10426 bp in length
            *    35274    35373: gap of unknown length
            *    35374    56929: contig of 21556 bp in length
            *    56930    57029: gap of unknown length
            *    57030    73825: contig of 16796 bp in length
            *    73826    73925: gap of unknown length
            *    73926    89138: contig of 15213 bp in length
            *    89139    89238: gap of unknown length
            *    89239   117743: contig of 28505 bp in length
            *   117744   117843: gap of unknown length
            *   117844   172622: contig of 54779 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172622
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-403E9"
     misc_feature    1. .6930
                     /note="assembly_name:Contig10"
     misc_feature    7031. .24747
                     /note="assembly_name:Contig11"
     misc_feature    24848. .35273
                     /note="assembly_name:Contig12"
     misc_feature    35374. .56929
                     /note="assembly_name:Contig13
                     clone_end:SP6
                     vector_side:right"
     misc_feature    57030. .73825
                     /note="assembly_name:Contig14"
     misc_feature    73926. .89138
                     /note="assembly_name:Contig15"
     misc_feature    89239. .117743
                     /note="assembly_name:Contig16"
     misc_feature    117844. .172622
                     /note="assembly_name:Contig17
                     clone_end:T7
                     vector_side:right"
BASE COUNT    49385 a   36666 c   36708 g   49156 t     707 others
ORIGIN


  Query Match             37.1%;  Score 284;  DB 2;  Length 172622;
  Best Local Similarity   91.2%;  Pred. No. 1.1e-72;
  Matches  299;  Conservative    2;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      12 aagatttcttctcttagtaattttatttctgccacgtgagatgacaagttctgttttaac 71
           | | | | | ||||   ||  |  ||   | |  | |    |  || ||||||||||||||
Db 160044 AGGTTGTTTACTCTGTATATCTAAATAGTTTCTCCCTATACTCCCAGGTTCTGTTTTAAC 159985

Qy      72 tgtgaatggtaaaactgagaactatatcctggatactacacctggctcccaagcatctct 131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159984 TGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCTGGCTCCCAAGCATCTCT 159925

Qy     132 gatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggg 191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159924 GATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTCTGGTACCGAGAGGAGGG 159865

Qy     192 gagagtggatttgaaatctgsaaacaaatcaattccagctctgtctgtgtctcttccat 251
           ||||||||||||||||||:||||||:||||||||||||||||||||||||||||||||||
Db 159864 GAGAGTGGATTTGAAATCTGGAAACAAAATCAATTCCAGCTCTGTCTGTGTCTCTTCCAT 159805

Qy     252 cagtgaaaatgacaacggaatcagctttacctgcaggctggggagggatcagtccgtgtc 311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159804 CAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGGAGGGATCAGTCCGTGTC 159745

Qy     312 cgtttcggtggtgctgaatgttacttttt 339
           ||||||||||||||||||||||||||| |
Db 159744 CGTTTCGGTGGTGCTGAATGTTACTTGT 159717


RESULT   6
AX245713
LOCUS       AX245713                338 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 643 from Patent WO0166753.
ACCESSION   AX245713
VERSION     AX245713.1  GI:15860387
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 643 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       93 a     69 c     91 g     85 t
ORIGIN


  Query Match             36.3%;  Score 277.4;  DB 6;  Length 338;
  Best Local Similarity   98.2%;  Pred. No. 7e-71;
  Matches  278;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      55 acaagttctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacct 114
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 AGAGGTTCTGTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCT 115

Qy     115 ggctcccaagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctc 174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 GGCTCCCAAGCATCTCTGATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTC 175

Qy     175 tggtaccgagaggaggggagagtggatttgaaaatctgsaaacaaatcaattccagctct 234
           ||||||||||||||||||||||||||||||||||||:||||||:||||||||||||||||
Db     176 TGGTACCGAGAGGAGGGGAGAGTGGATTTGAAATCTGGAAACAAAATCAATTCCAGCTCT 235

Qy     235 gtctgtgtctcttccatcagtgaaaatgacaacggaatcagctttacctgcaggctgggg 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 GTCTGTGTCTCTTCCATCAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGG 295

Qy     295 agggatcagtccgtgtccgtttcggtggtgctgaatgttactt 337
           ||||||||||||||||||||||||| |||||||||||||||||
Db     296 AGGGATCAGTCCGTGTCCGTTTCGGAGGTGCTGAATGTTACTT 338


RESULT   7
AX246249
LOCUS       AX246249                361 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 1179 from Patent WO0166753.
ACCESSION   AX246249
VERSION     AX246249.1  GI:15860923
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 1179 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .361
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       114 a      63 c      90 g      94 t
ORIGIN


  Query Match             29.7%;  Score 227.2;  DB 6;  Length 361;
  Best Local Similarity   99.1%;  Pred. No. 4.8e-56;
  Matches  226;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgcaaaatgggaagatttcttctcttagtaatttttatttctgccacgtgaqatgacaaqt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 ATGCAAATGGGAAGATTTCTTCTCTTAGTAATTTTATTTCTGCCACGTGAGATGACAAGT 192

Qy      61 tctgttttaactgtgaatggtaaaactgaqaactatatcctggatactacacctggctcc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 TCTGTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTCCATACTACACCTGGCTCC 252

Qy     121 caaqcatctctqatatqtqctgttcaaaaccacaccagagaggaagaactgctctggtac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 CAAGCATCTCTGATATGTGCTGTTCAAAACCACACCACAGAGGAAGAACTGCTCTGGTAC 312

Qy     181 cqaqaqqaqqqqqaqaqtqqatttqaaatctqsaaacaaaatcaattcc 228
           |||||||||||||||||||||||||||||:|||||:|||||||||
Db     313 CGAGAGGAGGGGAGAGTGGATTTGAAATCTGGAAACAAAATCAATTCC 360


RESULT   8
AX246408
LOCUS       AX246408                 328 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 1338 from Patent WO0166753.
ACCESSION   AX246408
VERSION     AX246408.1  GI:15861082
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 1338 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1..328
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       102 a      62 c      81 g      83 t
ORIGIN


  Query Match             29.5%;  Score 225.4;  DB 6;  Length 328;
  Best Local Similarity   97.8%;  Pred. No. 1.6e-55;
  Matches  226;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 atgcaaatgggaagatttcttctcttagtaatttttatttctgccacgtgagatgacaaqt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97 ATGCAAATGGGAAGATTTCTTCTCTTAGTAATTTTATTTCTGCCACGTGAGATGACAAGT 156

Qy      61 tctgttttaactgtgaatggtaaaactgaqaactatatcctggatactacacctggctcc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     157 TCTGTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCTGGCTCC 216

Qy     121 caaqcatctctqatatqtqctgttcaaaaccacaccagaqaggaagaactgctctggtac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     217 CAAGCATCTCTGATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTCTGGTAC 276

Qy     181 cqaqaqqaqqqqqaqaqtqqatttqaaatctqsaaacaaaatcaattccaqc 231
           ||| |||||||||||||||||||||||||||:|||||:||| ||||| ||
Db     277 CGACAGGAGGGGAGAGTGGATTTGAAATCTGGAAACAAAATCCATTCCCGC 327


RESULT   9
AX070657
LOCUS       AX070657                 359 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 1129 from Patent WO0102568.
ACCESSION   AX070657
VERSION     AX070657.1  GI:12580442
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
            Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Lamson,G., Drmanac,R., Crkvenjakov,R., Drmanac,S., Dickson,M.,
            Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and Stache Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0102568-A 1129 11-JAN-2001;
            CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1..359
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    1..359
                     /note="n = A,T,C or G"
BASE COUNT       113 a      64 c      92 g      89 t       1 others
ORIGIN


  Query Match             29.2%;  Score 223.6;  DB 6;  Length 359;
  Best Local Similarity   98.7%;  Pred. No. 5.6e-55;
  Matches  223;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 atgcaaatgggaagatttcttctcttagtaatttttatttctgccacgtgaqatgacaaqt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 ATGCAAATGGGAAGATTTCTTCTCTTAGTAATTTTATTTCTGCCACGTGAGATGACAAGT 192

Qy      61 tctgttttaactgtgaatggtaaaactgaqaactatatcctggatactacacctggctcc 120
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     193 TCTGTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCTGGCTCC 252

Qy     121 caaqcatctctqatatqtqctgttcaaaaccacaccagaqaggaagaactgctctggtac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 CAAGCATCTCTGATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTCTGGTAC 312

Qy     181 cqaqaqqaqqqqqaqaqtqqatttqaaatctqsaaacaaaatcaatt 226
           |||||||||||||||||||||||||||||||:|||||:|||||||
Db     313 CGACAGGAGGGGAGAGTGGATTTGAAATCTGGAAACAAAATCAATT 358


RESULT  10
AL645479
LOCUS       AL645479                201797 bp    DNA     linear   HTG 01-FEB-2002
DEFINITION  Mus musculus chromosome 11 clone RP23-39603, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL645479
VERSION     AL645479.7  GI:18491374
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (31 JAN 2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests, clonerequest@sanger.ac.uk
COMMENT     On Feb 4, 2002 this sequence version replaced gi:18476855.
            ---------- Genome Center
```

```
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM39603
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 196898 bases at least Q40
            Consensus quality: 198094 bases at least Q30
            Consensus quality: 198893 bases at least Q20
            Insert size: 200297; sum-of-contigs
            Insert size: 187260; 5.3% error; agarose-fp
            Quality coverage: 7.48x in Q20 bases; sum-of-contigs quality
            coverage: 8.00x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .201797
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-39603"
                     /clone_lib="RPCI-23"
     misc_feature    1. .11113
                     /note="assembly_fragment:02332
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    11214. .48611
                     /note="assembly_fragment:02747
                     fragment_chain:1"
     misc_feature    48712. .59723
                     /note="assembly_fragment:00080
                     fragment_chain:2"
     misc_feature    59824. .80048
                     /note="assembly_fragment:02274
                     fragment_chain:2"
     misc_feature    80149. .83718
                     /note="assembly_fragment:01802
                     fragment_chain:2"
     misc_feature    83819. .109110
                     /note="assembly_fragment:00974
                     fragment_chain:2"
     misc_feature    109211. .128137
                     /note="assembly_fragment:02689
                     fragment_chain:2"
     misc_feature    128238. .137020
                     /note="assembly_fragment:02530
                     fragment_chain:3"
     misc_feature    137121. .142890
                     /note="assembly_fragment:03193
                     fragment_chain:3"
     misc_feature    142991. .152615
                     /note="assembly_fragment:00674
                     fragment_chain:4"
     misc_feature    152716. .159209
                     /note="assembly_fragment:00421
                     fragment_chain:4"
     misc_feature    159310. .162315
                     /note="assembly_fragment:01093
                     fragment_chain:4"
     misc_feature    162416. .181977
                     /note="assembly_fragment:02178
                     fragment_chain:4"
     misc_feature    182078. .193642
                     /note="assembly_fragment:01392
                     fragment_chain:4"
     misc_feature    193743. .198053
                     /note="assembly_fragment:04059
                     fragment_chain:4"
     misc_feature    198154. .201797
                     /note="assembly_fragment:00302
                     fragment_chain:4
                     clone_end:T7
                     vector_side:right"
BASE COUNT    57371 a  42459 c  42767 g  57700 t   1500 others
ORIGIN


  Query Match             28.1%;  Score 214.6;  DB 2;  Length 201797;
  Best Local Similarity   82.1%;  Pred. No. 3.6e-52;
  Matches  247;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     326 tgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggca 385
           ||| | ||||  |||||||||| |||||||| || | ||||||||| || ||||||  ||
Db  123219 TGAGTCTTACAGTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTCGAGGAAAACA 123278

Qy     386 gtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtaca 445
           |  |||| |  |||||||||||| |||||| ||||||| || || ||||||||||||| |
Db  123279 GCGATGTAAGTTTGGTTTGCAACGTGAAATCCAAGCCTCAAGCGCAAATGATGTGGTATA 123338

Qy     446 aaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagt 505
           |||||| ||   | | | |||||||||| |||||||||||||| |||||||| ||||
Db  123339 AAAACAACAGCGCTTGGTTTTAGAGAAAGCCGTCACCAAATCCACCAGACAAGGGAGT 123398

Qy     506 cttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattg 565
           |||||||| |||| |||||||||||| |||| |||||| || |||||||| |||||||
Db  123399 CTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGCTGTATTG 123458

Qy     566 caaagtcatctctgaaaacggagagcttggactttcacctgattgttaaaqataaaactq 625
           | |  |||||||||||||| ||||| | |||||| |||||| |||||||||| |||  ||
Db  123459 CCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGGTAACTCTC 123518

Qy     626 t 626
           |
Db  123519 T 123519


RESULT  11
AL603842/c
LOCUS       AL603842              229289 bp    DNA     linear   ROD 17 NOV 2001
DEFINITION  Mouse DNA sequence from clone RP23-100P23 on chromosome 11,
            complete sequence.
ACCESSION   AL603842
VERSION     AL603842.9  GI:16519591
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 229289)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 30, 2001 this sequence version replaced gi:16304764.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
```

```
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-100P23 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP23-100P23.
FEATURES             Location/Qualifiers
     source          1. .229289
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-100P23"
                     /clone_lib="RPCI-23"
     misc_feature    59782. .59788
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    125096. .125397
                     /note="Single clone region. Sequence from clone PCR only."
     misc_feature    170770. .170774
                     /note="1327 bases of IS2 transposon (V00610) removed here.
                     This sequence represents the duplicated flanking sequence
                     of the IS2."
     misc_feature    176567. .176641
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    64835 a  50446 c  50226 g  63782 t
ORIGIN


  Query Match             28.1%;  Score 214.6;  DB 10;  Length 229289;
  Best Local Similarity   82.1%;  Pred. No. 4.6e-52;
  Matches  247;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     126 tgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggca 385
           ||| | ||||  |||||||||| |||||||| || | ||||||||| || ||||||  ||
Db  191296 TGAGTCTTACAGTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACA 191237

Qy     386 gtaatgtgaagttggtttgcaatgtgaaagccaaccccaggctcaaatgatgtggtaca 445
           | |||| |  |||||||||| |||||| ||||||| || || ||||||||||||| |
Db  191236 GCGATGTAAGTTTGGTTTGCAACGTCAAATCCAACCCTCAAGCGCAAATGATGTGGTATA 191177

Qy     446 aaaacaqtaqtctcctcqatttaqaqaaaaqccqtcaccaaatccaacagacaaqtgaqt 505
           ||||||  ||   | | | |||||||||| ||||||||||||||| ||||||| ||||
Db  191176 AAAACAACAGCGCCTTGGTTTTTAGAGAAAGGGGTTCACCAAATCCACCAGACAAGGGAGT 191117

Qy     506 ctttcaqctgtcaatcaccaaaqtcqaqaaqcctqacaacqaaacctacagttgtattg 565
           |||||||| |||| |||||||||||| |||  ||||||| || |||||||| |||||||
Db  191116 CTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGCTGTATTG 191057

Qy     566 caaaqtcatctctqaaaacggaqaqcttqqacttcacctqattgttaaagataaaactg 625
           | |  |||||||||||| ||||| | |||||| |||||| ||||||||| |||  ||
Db  191056 CCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGGTAACTCTC 190997

Qy     626 t 626
           |
Db  190996 T 190996


RESULT  12
AX070653
LOCUS       AX070653                359 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 1125 from Patent WO0102568.
ACCESSION   AX070653
VERSION     AX070653.1  GI:12580438
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          . Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
            Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Lamson,G., Drmanac,R., Crkenjakov,R., Drmanac,S., Dickson,M.,
            Labat,I., Leshkowitiz,D., Kita,D., Garcia,V. and Strache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0102568 A 1125 11 JAN 2001;
            CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .359
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    1. .359
                     /note="n = A,T,C or G"
BASE COUNT      106 a     69 c     95 g     88 t      1 others
ORIGIN


  Query Match             23.5%;  Score 179.6;  DB 6;  Length 359;
  Best Local Similarity   94.5%;  Pred. No. 4.9e-42;
  Matches  206;  Conservative    1;  Mismatches    9;  Indels    2;  Gaps    2;

Qy       1 atgcaaatgggaagatttcttctcttagtaatttatttctgccacgtgagatgacaagt 60
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db     133 ATGCAAATGGGAAGATTTCTTCTCTTAGTAATTCTATTTCTGCCACGTGAGATGACAAGT 192

Qy      61 tctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacctggctcc 120
           |||| |||||||||||||| |||| ||||||||||||||||||||||||||||||||||
Db     193 TCTGGTTTAACTGTGAATGGTAACACTGAGAACTATATCCTGGATACTACACCTGGCTCC 252

Qy     121 caagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtac 180
           |||||||||||||||||||||||||| |||||||||||||||||| ||||||||||||
Db     253 CAAGCATCTCTGATATGTGCTGTTCCAAACCACACCAGAGAGGAAG-ACTGCTCTGGTAC 311

Qy     181 cqaqaqqaqqqgagagtgqgatttqaaatctgsaaacaa 218
           |||||||| ||||||| |||||||| ||||||:|| ||
Db     312 CGAGAGGA-GGGAGAGGGGATTTGANATCTGGAACAAA 348


RESULT  13
AC103795
LOCUS       AC103795               70248 bp    DNA     linear   HTG 29-NOV-2001
DEFINITION  Homo sapiens chromosome 11 clone CTD 2529D2 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC103795
VERSION     AC103795.1  GI:17149680
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70248)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone CTD 2529D2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70248)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beguslavkiy,D., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand Pierre,N.,
            Haqos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Keils,C., Lakocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenqa,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
```

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L21770
Center clone name: 2529_D_2
---------------.
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 687: contig of 687 bp in length
* 688 787: gap of 100 bp
* 788 1501: contig of 714 bp in length
* 1502 1601: gap of 100 bp
* 1602 2345: contig of 744 bp in length
* 2346 2445: gap of 100 bp
* 2446 3151: contig of 706 bp in length
* 3152 3251: gap of 100 bp
* 3252 3929: contig of 678 bp in length
* 3930 4029: gap of 100 bp
* 4030 4719: contig of 690 bp in length
* 4720 4819: gap of 100 bp
* 4820 5538: contig of 719 bp in length
* 5539 5638: gap of 100 bp
* 5639 6356: contig of 718 bp in length
* 6357 6456: gap of 100 bp
* 6457 7156: contig of 700 bp in length
* 7157 7256: gap of 100 bp
* 7257 7888: contig of 632 bp in length
* 7889 7988: gap of 100 bp
* 7989 8743: contig of 755 bp in length
* 8744 8843: gap of 100 bp
* 8844 9533: contig of 690 bp in length
* 9534 9633: gap of 100 bp
* 9634 10371: contig of 738 bp in length
* 10372 10471: gap of 100 bp
* 10472 11203: contig of 732 bp in length
* 11204 11303: gap of 100 bp
* 11304 12016: contig of 713 bp in length
* 12017 12116: gap of 100 bp
* 12117 12823: contig of 707 bp in length
* 12824 12923: gap of 100 bp
* 12924 13628: contig of 705 bp in length
* 13629 13728: gap of 100 bp
* 13729 14441: contig of 713 bp in length
* 14442 14541: gap of 100 bp
* 14542 15272: contig of 731 bp in length
* 15273 15372: gap of 100 bp
* 15373 16106: contig of 734 bp in length
* 16107 16206: gap of 100 bp
* 16207 16935: contig of 729 bp in length
* 16936 17035: gap of 100 bp
* 17036 17762: contig of 727 bp in length
* 17763 17862: gap of 100 bp
* 17863 18558: contig of 696 bp in length
* 18559 18658: gap of 100 bp
* 18659 19351: contig of 693 bp in length
* 19352 19451: gap of 100 bp
* 19452 20181: contig of 730 bp in length
* 20182 20281: gap of 100 bp
* 20282 21023: contig of 742 bp in length
* 21024 21123: gap of 100 bp
* 21124 21827: contig of 704 bp in length
* 21828 21927: gap of 100 bp
* 21928 22624: contig of 697 bp in length
* 22625 22724: gap of 100 bp
* 22725 23401: contig of 677 bp in length
* 23402 23501: gap of 100 bp
* 23502 24228: contig of 727 bp in length
* 24229 24328: gap of 100 bp
* 24329 25055: contig of 727 bp in length
* 25056 25155: gap of 100 bp
* 25156 25898: contig of 743 bp in length
* 25899 25998: gap of 100 bp
* 25999 26710: contig of 712 bp in length
* 26711 26810: gap of 100 bp
* 26811 27551: contig of 741 bp in length
* 27552 27651: gap of 100 bp
* 27652 28394: contig of 743 bp in length
* 28395 28494: gap of 100 bp
* 28495 29230: contig of 736 bp in length
* 29231 29330: gap of 100 bp
* 29331 30037: contig of 707 bp in length
* 30038 30137: gap of 100 bp
* 30138 30854: contig of 717 bp in length
* 30855 30954: gap of 100 bp
* 30955 31680: contig of 726 bp in length
* 31681 31780: gap of 100 bp
* 31781 32520: contig of 740 bp in length
* 32521 32620: gap of 100 bp
* 32621 33354: contig of 734 bp in length
* 33355 33454: gap of 100 bp
* 33455 34191: contig of 737 bp in length
* 34192 34291: gap of 100 bp
* 34292 35034: contig of 743 bp in length
* 35035 35134: gap of 100 bp
* 35135 35880: contig of 746 bp in length
* 35881 35980: gap of 100 bp
* 35981 36663: contig of 683 bp in length
* 36664 36763: gap of 100 bp
* 36764 37477: contig of 714 bp in length
* 37478 37577: gap of 100 bp
* 37578 38321: contig of 744 bp in length
* 38322 38421: gap of 100 bp
* 38422 39110: contig of 689 bp in length
* 39111 39210: gap of 100 bp
* 39211 39892: contig of 682 bp in length
* 39893 39992: gap of 100 bp
* 39993 40678: contig of 686 bp in length
* 40679 40778: gap of 100 bp
* 40779 41479: contig of 701 bp in length
* 41480 41579: gap of 100 bp
* 41580 42310: contig of 731 bp in length
* 42311 42410: gap of 100 bp
* 42411 43144: contig of 734 bp in length
* 43145 43244: gap of 100 bp
* 43245 43985: contig of 741 bp in length
* 43986 44085: gap of 100 bp
* 44086 44777: contig of 692 bp in length
* 44778 44877: gap of 100 bp
* 44878 45598: contig of 721 bp in length
* 45599 45698: gap of 100 bp
* 45699 46433: contig of 735 bp in length
* 46434 46533: gap of 100 bp
* 46534 47267: contig of 734 bp in length

```
              *    47268 47367: gap of      100 bp
              *    47368    48058: contig of 691 bp in length
              *    48059 48158: gap of      100 bp
              *    48159    48880: contig of 722 bp in length
              *    48881 48980: gap of      100 bp
              *    48981    49717: contig of 737 bp in length
              *    49718 49817: gap of      100 bp
              *    49818    50560: contig of 743 bp in length
              *    50561 50660: gap of      100 bp
              *    50661    51400: contig of 740 bp in length
              *    51401 51500: gap of      100 bp
              *    51501    52227: contig of 727 bp in length
              *    52228 52327: gap of      100 bp
              *    52328    53066: contig of 739 bp in length
              *    53067 53166: gap of      100 bp
              *    53167    53854: contig of 687 bp in length
              *    53855 53953: gap of      100 bp
              *    53954    54675: contig of 722 bp in length
              *    54676 54775: gap of      100 bp
              *    54776    55508: contig of 733 bp in length
              *    55509 55608: gap of      100 bp
              *    55609    56347: contig of 739 bp in length

  Query Match             14.3%;  Score 109.6;  DB 2;  Length 70248;
  Best Local Similarity   92.7%;  Pred. No. 1.6e-21;
  Matches  115;  Conservative    9;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      608 ttqttaaaqataaaactqtqqqtqtaccaataqaqcccattattgctgcatqtgttqtga 667
            ||  || |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    44391 TTCCTACAGATAAAACTGTGGGTGTACCAATAGAGCCCATTATTGCTGCATGTGTTGTGA 44450

Qy      668 tctttctqacattqtqcttqqactqattqctaqaaqaaaqaaaaataatgaaqctctgca 727
            |||||||||||||||||||||||||||||||||||||||||||||||||| | |  |
Db    44451 TCTTTCTGACATTGTGCTTTGGACTGATTGCTAGAAGAAAGAAAATAATGAAGGTATCTA 44510

Qy      728 tgaa 741
            | |
Db    44511 AGTA 44514


RESULT  14
AC087440/c
LOCUS       AC087440               217730 bp    DNA     linear   HTG 16-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11 425G13 map 11, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC087440
VERSION     AC087440.2  GI:13357819
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 217730)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-425G13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 217730)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeAtellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Glinde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Haqos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Leboczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Menenes,I., Mihova,T., Mlenqa,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schaaer,S., Schupback,R., Seaman,S., Severy,P.,
            Souqnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Travers,M., Travis,N., Triqilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 15, 2001 this sequence version replaced gi:12039453.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11822
                Center clone name: 425_G_13
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 192528 bases at least Q40
                Consensus quality: 205654 bases at least Q30
                Consensus quality: 210625 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 215130; sum-of-contigs
                Quality coverage: 4.0 in Q20 bases; agarose-fp
                Quality coverage: 3.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   182949: contig of 182949 bp in length
            *  182950 183049: gap of      100 bp
            *  183050   183137: contig of 88 bp in length
            *  183138 183237: gap of      100 bp
            *  183238   183640: contig of 403 bp in length
            *  183641 183740: gap of      100 bp
            *  183741   184713: contig of 973 bp in length
            *  184714 184813: gap of      100 bp
            *  184814   185319: contig of 506 bp in length
            *  185320 185419: gap of      100 bp
            *  185420   186040: contig of 621 bp in length
            *  186041 186140: gap of      100 bp
            *  186141   186820: contig of 680 bp in length
            *  186821 186920: gap of      100 bp
            *  186921   187855: contig of 935 bp in length
            *  187856 187955: gap of      100 bp
            *  187956   188593: contig of 638 bp in length
            *  188594 188693: gap of      100 bp
            *  188694   189289: contig of 596 bp in length
            *  189290 189389: gap of      100 bp
            *  189390   190822: contig of 1433 bp in length
            *  190823 190922: gap of      100 bp
            *  190923   191663: contig of 741 bp in length
            *  191664 191763: gap of      100 bp
            *  191764   192866: contig of 1103 bp in length
            *  192867 192966: gap of      100 bp
            *  192967   194120: contig of 1154 bp in length
            *  194121 194220: gap of      100 bp
            *  194221   194904: contig of 684 bp in length
            *  194905 195004: gap of      100 bp
            *  195005   195907: contig of 903 bp in length
            *  195908 196007: gap of      100 bp
            *  196008   197287: contig of 1280 bp in length
```

```
            *   197288 197387: gap of      100 bp
            *   197388   198887: contig of 1500 bp in length
            *   198888 198987: gap of      100 bp
            *   198988   200386: contig of 1399 bp in length
            *   200387 200486: gap of      100 bp
            *   200487   201763: contig of 1277 bp in length
            *   201764 201863: gap of      100 bp
            *   201864   203900: contig of 2037 bp in length
            *   203901 204000: gap of      100 bp
            *   204001   205876: contig of 1876 bp in length
            *   205877 205976: gap of      100 bp
            *   205977   207807: contig of 1831 bp in length
            *   207808 207907: gap of      100 bp
            *   207908   210489: contig of 2582 bp in length
            *   210490 210589: gap of      100 bp
            *   210590   213765: contig of 3176 bp in length
            *   213766 213865: gap of      100 bp
            *   213866   217087: contig of 3222 bp in length
            *   217088 217187: gap of      100 bp
            *   217188   217730: contig of 543 bp in length.
FEATURES             Location/Qualifiers
     source          1. .217730
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-425G13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .182949
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    183050. .183137
                     /note="assembly_fragment"
     misc_feature    183238. .183640
                     /note="assembly_fragment"
     misc_feature    183741. .184713
                     /note="assembly_fragment"
     misc_feature    184814. .185319
                     /note="assembly_fragment"
     misc_feature    185420. .186040
                     /note="assembly_fragment"
     misc_feature    186141. .186820
                     /note="assembly_fragment"
     misc_feature    186921. .187855
                     /note="assembly_fragment"
     misc_feature    187956. .188593
                     /note="assembly_fragment"
     misc_feature    188694. .189289
                     /note="assembly_fragment"
     misc_feature    189390. .190822
                     /note="assembly_fragment"
     misc_feature    190923. .191663
                     /note="assembly_fragment"
     misc_feature    191764. .192866
                     /note="assembly_fragment"
     misc_feature    192967. .194120
                     /note="assembly_fragment"
     misc_feature    194221. .194904
                     /note="assembly_fragment"
     misc_feature    195005. .195907
                     /note="assembly_fragment"
     misc_feature    196008. .197287
                     /note="assembly_fragment"
     misc_feature    197388. .198887
                     /note="assembly_fragment"
     misc_feature    198988. .200386
                     /note="assembly_fragment"
     misc_feature    200487. .201763
                     /note="assembly_fragment"
     misc_feature    201864. .203900
                     /note="assembly_fragment"
     misc_feature    204001. .205876
                     /note="assembly_fragment"
     misc_feature    205977. .207807
                     /note="assembly_fragment"
     misc_feature    207908. .210489
                     /note="assembly_fragment"
     misc_feature    210590. .213765
                     /note="assembly_fragment"
     misc_feature    213866. .217087
                     /note="assembly_fragment"
     misc_feature    217188. .217730
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    69080 a  40450 c  39974 g  65624 t   2602 others
ORIGIN


  Query Match             5.7%;  Score 43.8;  DB 2;  Length 217730;
  Best Local Similarity   95.7%;  Pred. No. 0.11;
  Matches   45;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     719 agctctgcatgaaggataaagaccctcacagtgaaacagctctatga 765
           ||||||||||||||||||||||||||||||||||||||||||||| | |
Db 199852 AGCTCTGCATGAAGGATAAAGACCCTCACAGTGAAACAGCTCTGTAA 199806


RESULT  15
AF429315/c
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
```

```
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGRSYCGGWFDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a   32731 c   30696 g   28283 t    4254 others
ORIGIN


  Query Match             5.5%;  Score 42;  DB 9;  Length 125020;
  Best Local Similarity   8.6%;  Pred. No. 0.35;
  Matches   36;  Conservative  211;  Mismatches  170;  Indels    0;  Gaps    0;

Qy     142 qttcaaaaccacaccaqaqaqqaagaaactqctctqqtaccgaqaqqqaggqqqagagtggat 201
           ::::::|:  ::|:::    |  | ::::  |::  |::::::: :::|  :    : ::
Db   61034 DKDHHAWSRSRCVHSGASMACWATVRVHGWCWVGWGDDMMSWSCRKMGWMKCMSHDWRDR 60975

Qy     202 ttgaaattctqaaaacaamatcaattccaqctctqtctqtqtctcttccatcaqtqaaaat 261
           ::::::  :|: :: : : ::::::    :: : ::::: :::::::: :::: : :  ::
Db   60974 HHDVWMCAKGRCWDGMBWSYSWMWDGWBHMKWDHBBHDBWYSWVWDBDRDBWMWMWCBRD 60915

Qy     262 qacaacqgaatcaqctttaccctgcaggctqqggagggatcaqtccgtgtccgtttcggtg 321
           :: : ::  :::  :: :    ::  :::: : :: :: :::  ::   : :
Db   60914 RDDHYYDYSDBBBWHYVYYWKNHBWYHHWSSWYDYHSWSHHNYHVWRMHVRMVBVHHNMN 60855

Qy     322 qtqctqaatqttactttttcctcctctcctaagtqqaaacgactttccaaacagttgaqgaa 381
           :: ::::    ::   : : ::  :::::: : : ::: : ::   ||:: :|    :
Db   60854 MWVDWSMMMMVNMHVRVBABMVMMNBSMWVVMWNDNMRYWMAHKRGCAAMMMKTMSACMT 60795

Qy     382 qqcaqtaatqtqaaqttqqtttqcaatgtqaaaqccaacccccaggctcaaatqatqtgg 441
             : : ::|    :: :  :     ::  |: ::|::| :  ::::   :|: |: : ::
Db   60794 TMSCRSPWTCAACWRAKWMKMMYRCPWMACMCWSCMRARYWKMWRRGAWWAMNCPPKPSS 60735

Qy     442 tacaaaaacaqtaqtctcctcgatttaqaqaaaaqccqtcaccaaatccaacaqacaagt 501
           :| :::::| : : ::  : : :  : :: :|::     :: :::: :| |: : :::
Db   60734 CMCCRMMRMAMWGRRMWWAWWKYWRMMWWKYWARKAGARAMSRRRWYTMACCMWMWWRRC 60675

Qy     502 qaqtcttttcaqctqtcaatcaccaaaqtcqaqaaqcctgacaacggaacctacagt 558
           :: |:  :|:::      :|:|:::|: ::|   :|:: :: ::|:  || | |||
Db   60674 MRRRCYGRWCWSSWACWYCMSAMCWMARRMSAWYYRCSYAMMYMMGSCTCCCAAAGT 60618


Search completed: September 17, 2002, 09:59:08
Job time: 5856 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 09:59:08 ; Search time 2925.1 Seconds
(without alignments)
5644.606 Million cell updates/sec

Title: US-09-863-823-5
Perfect score: 789
Sequence: 1 atggcatggaagagcagtgt acagtgaaaragctctatga 789

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	504.6	64.0	1340	10	BC022683	BC022683 Mus muscu
c 2	286.6	36.3	70248	2	AC103795	AC103795 Homo sapi
c 3	286.6	36.3	171136	9	AC090685	AC090685 Homo sapi
c 4	286.6	36.3	172046	2	AC087644	AC087644 Homo sapi
c 5	284	36.0	172622	2	AC024065	AC024065 Homo sapi
6	277.4	35.2	338	6	AX245713	AX245713 Sequence
7	250.8	31.8	361	6	AX246249	AX246249 Sequence
8	249	31.6	328	6	AX246408	AX246408 Sequence
9	247.2	31.3	359	6	AX070657	AX070657 Sequence
10	214.6	27.2	201797	2	AL645479	AL645479 Mus muscu
c 11	214.6	27.2	229289	10	AL603842	AL603842 Mouse DNA
12	203.4	25.8	359	6	AX070653	AX070653 Sequence
13	109.6	13.9	70248	2	AC103795	AC103795 Homo sapi
c 14	43.8	5.6	217730	2	AC087440	AC087440 Homo sapi
c 15	42	5.3	125020	9	AF429315	AF429315 Homo sapi
16	42	5.3	217730	2	AC087440	AC087440 Homo sapi
c 17	40.4	5.1	174493	9	AC025522	AC025522 Homo sapi
c 18	40.4	5.1	189126	2	AC021132	AC021132 Homo sapi
c 19	40.2	5.1	393	11	G37560	G37560 SHGC-57994
20	39.6	5.0	2134	3	DMU51473	U51473 Drosophila
21	39.6	5.0	8500	2	AC020364	AC020364 Drosophil
22	39.6	5.0	68727	3	AC004516	AC004516 Drosophil
c 23	39.6	5.0	193262	3	AC007579	AC007579 Drosophil
c 24	39.6	5.0	262173	3	AE003808	AE003808 Drosophil
25	39.2	5.0	2799	9	HSNCAME	X16841 Human mRNA
26	39.2	5.0	2960	9	S71824	S71824 N-CAM=145 k
27	39.2	5.0	3309	9	AA057509	AA057509 Homo sapi
28	39.2	5.0	4092	6	AR062746	AR062746 Sequence
29	39.2	5.0	4092	9	HUMROR2A	M97639 Human trans
30	39.2	5.0	8513	6	AX207284	AX207284 Sequence
31	39	4.9	172567	2	AC015493	AC015493 Homo sapi
c 32	39	4.9	203252	2	AC025421	AC025421 Homo sapi
33	38.8	4.9	157756	2	AP001075	AP001075 Homo sapi
c 34	38.8	4.9	195622	2	AP000807	AP000807 Homo sapi
35	38.6	4.9	2855	10	RNU34985	U34985 Rattus norv
36	38.6	4.9	2869	6	A72401	A72401 Sequence 2
37	38.6	4.9	2869	6	AR043363	AR043363 Sequence
38	38.6	4.9	2869	6	I60497	I60497 Sequence 2
c 39	38.6	4.9	65119	9	AL590062	AL590062 Human DNA
c 40	38.6	4.9	125698	9	HS692C8	AL034561 Human DNA
c 41	38.6	4.9	159287	9	AP003694	AP003694 Homo sapi
42	38.6	4.9	175218	2	AC027031	AC027031 Homo sapi
c 43	38.6	4.9	177731	2	AC018937	AC018937 Homo sapi
c 44	38	4.8	291	1	MC385	Z33288 M.capricolu
45	38	4.8	71023	2	AC004426	AC004426 Drosophil

ALIGNMENTS

RESULT 1
BC022683
LOCUS BC022683 1340 bp mRNA linear BOD 04-FEB-2002
DEFINITION Mus musculus, RIKEN cDNA 2010002A20 gene, clone IMAGE:4241194, mRNA, partial cds.
ACCESSION BC022683
VERSION BC022683.1 GI:18490500
KEYWORDS .
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1340)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-FEB 2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) med@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 43 Row: o Column: 9.
FEATURES              Location/Qualifiers
     source           1. .1340
                      /organism="Mus musculus"
                      /db_xref="LocusID:66601"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:4241194"
                      /tissue_type="Kidney, normal. 5 month old male mouse."
                      /clone_lib="NCI_CGAP_Kid14"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     CDS              <1  .776
                      /codon_start=3
                      /product="RIKEN cDNA 2010002A20 gene"
                      /protein_id="AAH22683.1"
                      /db_xref="GI:18490501"
                      /translation="ITGPLQACQLLLVVLSLPQGRTSSVLTVNGRTENYILDTQHGVQ
                      ASLECAVQNHTEDEELLWYREDGIVDLKNGNKINISSVCVSPINESDNGVRFTCKLQR
                      DQTVSVTVVLNVTFPPLLSGNGFQTVEENSDVSLVCNVKSNPQAQMMWYKNNSALVLE
                      KGRHQIHQTRESFQLSTTKVKKSDNGIYSCIASSSLKMEIMDFHLLVKDKVEVMPAEP
                      IIAACVVVVLTMAFALFSRRKRIMKLCGKKNDPNSETAL"
BASE COUNT       418 a     261 c     288 g     373 t
ORIGIN


  Query Match             64.0%;  Score 594.6;  DB 10;  Length 1340;
  Best Local Similarity   79.5%;  Pred. No. 3.1e-137;
  Matches  594;  Conservative    2,  Mismatches  151,  Indels    0,  Gaps    0;

Qy      44 cttctcttaqtaattttatttctqccacqtqaqatqacaaqttctgttttaactgtqaat 102
           ||||||||||  |||||| | ||||||  | ||  ||||| ||||||||||||||||||
Db      30 CTTCTCTTAGTGGTTTTATCTTTGCCACAGGGCACAACAACCTCTGTTTTAACTGTCAAT 89

Qy     103 qqtaaaactqaqaactatatcctqqatactacacctqqctcccaagcatctctqatatqt 162
           |||| ||||||||||||||| ||||| ||| || ||| ||||||||||||| ||||
Db      90 GGTAGAACTGAGAACTATATTCTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGT 149

Qy     163 qctqttcaaaaccacacccaqaqqaqqaaqaactgctctqqtacaqagaggaqqqqqaqagtg 222
           ||||||||||||||||||   || |||||||| |||||||| |||| || || | |||
Db     150 GCTGTTCAAAACCACACCGAGGATGAAGAACTTCTCTGGTACAGAGAACATGGAATAGTA 209

Qy     223 qatttqaaatctqsaaacaamatcaattccaqctctqtctgtqtctcttccatcagtqaa 282
           ||| |||||  ||:|||||:|||||   ||| |||| |||||||| |||||| |||
Db     210 GATCTGAAAAATGCAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACCAA 269

Qy     283 aatqacaacqqaatcaqtttacctqcaqqctqqqqaqqqatcaqtccqtqtccqtttcq 342
           | |||||| ||| || ||||||||||| |||   | | |||||| | |||||||| |
Db     270 AGTGACAATGGAGTCCGCTTTACCTGCAAGCTACAGCGCGATCAGACGGTGTCCGTTACA 329

Qy     343 qtqqtqctqaatqttacttttcctcctctcctaaqtqqaaacqacttccaaacagttgag 402
           || |||||||| |||||  |||||||||||| |||||||| || | |||||||| || |||
Db     330 GTAGTGCTGAACGTTACCTTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGCAG 389

Qy     403 qaaqqcaqtaatqtqaaqttqqtttqcaatqtqaaaqccaacccccagqctcaaatqatq 462
           ||| ||| |||| | |||||||||| |||||| ||||||| || || |||||||||
Db     390 GAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATG 449

Qy     463 tqqtacaaaaacaqtaqtctcctcqatttagaqaaaaqccqtcaccaaatccaacaqaca 522
           ||||| |||||||  ||   | | | |||||||||| ||||||||||||||| ||||||
Db     450 TGGTATAAAAACAACAGCGCCTTGGTTTTAGAGAAAGCGCGTCACCAAATCCACCAGACA 509

Qy     524 aqtqaqtcttttcaqctqtcaatcaccaaaqtcqagaaqcctqacaacqqaacctacagt 582
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     510 AGTGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGC 569

Qy     583 tqtattqcaaaqtcatctctqaaaacqgagaqcttqqactttcacctqattqttaaaqat 642
           |||||||| |  |||||||||||||| ||||| | ||||||| |||||| ||||||||||
Db     570 TGTATTGCCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGAC 629

Qy     643 aaaactgtqqqtqtaccaatagaqcccattattgctgcatqtqttqtqatctttctqaca 702
           |||  | | |   | ||   |||||||| |||||||| ||||| ||| || | |||||
Db     630 AAAGTTTTTGTCATGCCGCCAGACCCCATCATTGCTGCTTGTGTCGTGGTCGTCCTGACG 689

Qy     703 ttqtgctttqqactqattgctagaagaaaqaaaataatgaagctctgcatqaaqgataaa 762
           ||   |||| ||| || |||| || || | ||||||||||||||   ||| | ||
Db     690 ATGGCTTTTGCCCTGTTTTCTAGGAGCAAAACAATAATGAAGCTCTGTGGGAAAAAGAAT 749

Qy     763 qaccctcacaqtqaaacagctctatqa 789
           ||||| |    ||||||||||||| |
Db     750 GACCCGAATTCAGAAACAGCTCTATAA 776


RESULT   2
AC103795/c
LOCUS       AC103795               70248 bp    DNA     linear   HTG 29-NOV-2001
DEFINITION  Homo sapiens chromosome 11 clone CTD-2529D2 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC103795
VERSION     AC103795.1  GI:17149680
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70248)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone CTD-2529D2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70248)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beguslavskiy,I., Boukhgalter,B.,
            Brown,A., Camarata,J., Campoplano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,F., Pieback,M., Riley,P., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21770
                Center clone name: 2529_D_2
```

```
--------------.
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      687: contig of 687 bp in length
*      688 787: gap of      100 bp
*      788     1501: contig of 714 bp in length
*     1502 1601: gap of      100 bp
*     1602     2345: contig of 744 bp in length
*     2346 2445: gap of      100 bp
*     2446     3151: contig of 706 bp in length
*     3152 3251: gap of      100 bp
*     3252     3929: contig of 678 bp in length
*     3930 4029: gap of      100 bp
*     4030     4719: contig of 690 bp in length
*     4720 4819: gap of      100 bp
*     4820     5538: contig of 719 bp in length
*     5539 5638: gap of      100 bp
*     5639     6356: contig of 718 bp in length
*     6357 6456: gap of      100 bp
*     6457     7156: contig of 700 bp in length
*     7157 7256: gap of      100 bp
*     7257     7888: contig of 632 bp in length
*     7889 7988: gap of      100 bp
*     7989     8743: contig of 755 bp in length
*     8744 8843: gap of      100 bp
*     8844     9533: contig of 690 bp in length
*     9534 9633: gap of      100 bp
*     9634    10371: contig of 738 bp in length
*    10372 10471: gap of      100 bp
*    10472    11203: contig of 732 bp in length
*    11204 11303: gap of      100 bp
*    11304    12016: contig of 713 bp in length
*    12017 12116: gap of      100 bp
*    12117    12823: contig of 707 bp in length
*    12824 12923: gap of      100 bp
*    12924    13628: contig of 705 bp in length
*    13629 13728: gap of      100 bp
*    13729    14441: contig of 713 bp in length
*    14442 14541: gap of      100 bp
*    14542    15272: contig of 731 bp in length
*    15273 15372: gap of      100 bp
*    15373    16106: contig of 734 bp in length
*    16107 16206: gap of      100 bp
*    16207    16935: contig of 729 bp in length
*    16936 17035: gap of      100 bp
*    17036    17762: contig of 727 bp in length
*    17763 17862: gap of      100 bp
*    17863    18558: contig of 696 bp in length
*    18559 18658: gap of      100 bp
*    18659    19351: contig of 693 bp in length
*    19352 19451: gap of      100 bp
*    19452    20181: contig of 730 bp in length
*    20182 20281: gap of      100 bp
*    20282    21023: contig of 742 bp in length
*    21024 21123: gap of      100 bp
*    21124    21827: contig of 704 bp in length
*    21828 21927: gap of      100 bp
*    21928    22624: contig of 697 bp in length
*    22625 22724: gap of      100 bp
*    22725    23401: contig of 677 bp in length
*    23402 23501: gap of      100 bp
*    23502    24228: contig of 727 bp in length
*    24229 24328: gap of      100 bp
*    24329    25055: contig of 727 bp in length
*    25056 25155: gap of      100 bp
*    25156    25898: contig of 743 bp in length
*    25899 25998: gap of      100 bp
*    25999    26710: contig of 712 bp in length
*    26711 26810: gap of      100 bp
*    26811    27551: contig of 741 bp in length
*    27552 27651: gap of      100 bp
*    27652    28394: contig of 743 bp in length
*    28395 28494: gap of      100 bp
*    28495    29230: contig of 736 bp in length
*    29231 29330: gap of      100 bp
*    29331    30037: contig of 707 bp in length
*    30038 30137: gap of      100 bp
*    30138    30854: contig of 717 bp in length
*    30855 30954: gap of      100 bp
*    30955    31680: contig of 726 bp in length
*    31681 31780: gap of      100 bp
*    31781    32520: contig of 740 bp in length
*    32521 32620: gap of      100 bp
*    32621    33354: contig of 734 bp in length
*    33355 33454: gap of      100 bp
*    33455    34191: contig of 737 bp in length
*    34192 34291: gap of      100 bp
*    34292    35034: contig of 743 bp in length
*    35035 35134: gap of      100 bp
*    35135    35880: contig of 746 bp in length
*    35881 35980: gap of      100 bp
*    35981    36663: contig of 683 bp in length
*    36664 36763: gap of      100 bp
*    36764    37477: contig of 714 bp in length
*    37478 37577: gap of      100 bp
*    37578    38321: contig of 744 bp in length
*    38322 38421: gap of      100 bp
*    38422    39110: contig of 689 bp in length
*    39111 39210: gap of      100 bp
*    39211    39892: contig of 682 bp in length
*    39893 39992: gap of      100 bp
*    39993    40678: contig of 686 bp in length
*    40679 40778: gap of      100 bp
*    40779    41479: contig of 701 bp in length
*    41480 41579: gap of      100 bp
*    41580    42310: contig of 731 bp in length
*    42311 42410: gap of      100 bp
*    42411    43144: contig of 734 bp in length
*    43145 43244: gap of      100 bp
*    43245    43985: contig of 741 bp in length
*    43986 44085: gap of      100 bp
*    44086    44777: contig of 692 bp in length
*    44778 44877: gap of      100 bp
*    44878    45598: contig of 721 bp in length
*    45599 45698: gap of      100 bp
*    45699    46433: contig of 735 bp in length
*    46434 46533: gap of      100 bp
*    46534    47267: contig of 734 bp in length
*    47268 47367: gap of      100 bp
*    47368    48058: contig of 691 bp in length
*    48059 48158: gap of      100 bp
*    48159    48880: contig of 722 bp in length
*    48881 48980: gap of      100 bp
*    48981    49717: contig of 737 bp in length
*    49718 49817: gap of      100 bp
*    49818    50560: contig of 743 bp in length
*    50561 50660: gap of      100 bp
*    50661    51400: contig of 740 bp in length
*    51401 51500: gap of      100 bp
*    51501    52227: contig of 727 bp in length
*    52228 52327: gap of      100 bp
*    52328    53066: contig of 739 bp in length
*    53067 53166: gap of      100 bp
*    53167    53853: contig of 687 bp in length
*    53854 53953: gap of      100 bp
*    53954    54675: contig of 722 bp in length
*    54676 54775: gap of      100 bp
```

```
            *     54776     55508: contig of 733 bp in length
            *     55509 55608: gap of      100 bp
            *     55609     56347: contig of 739 bp in length

  Query Match             36.3%;  Score 286.6;  DB 2;  Length 70248;
  Best Local Similarity   97.0%;  Pred. No. 4.6e-73;
  Matches  292;  Conservative    0;  Mismatches    9;  Indels    0,  Gaps    0,

Qy      350 tqaatqttactttttctctctctcctaaqtqqaaacqacttccaaaacagttqaqqaaqqca 409
            ||| | || |  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49441 TGAGTCTTGCAGTTCCTCCTCTCCTAAGTGGAAACGACTTCCAAACAGTTGAGGAAGGCA 49372

Qy      410 qtaatqtqaaqttqqtttqcaatqtqaaaqccaaccccccaqqctcaaaatqatqtqqtaca 469
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49371 GTAATGTGAAGTTGGTTTGCAATGTGAAAGCCAACCCCCAGGCTCAAATGATGTGGTACA 49312

Qy      470 aaaacaqtaqtctcctcqatttaqaqaaaaqccqtcaccaaatccaacaqacaaqtqaqt 529
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49311 AAAACAGTAGTCTCCTCGATTTAGAGAAAAGCCGTCACCAAATCCAACAGACAAGTGAGT 49252

Qy      530 cttttcaqctqtcaatcaccaaaqtcqaqaaqcctqacaacqqaacctacaqttqtatta 589
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49251 CTTTTCAGCTGTCAATCACCAAAGTCGAGAAGCCTGACAACGGAACCTACAGTTGTATTG 49192

Qy      590 caaaqtcatctctqaaaacqqaqaqctttqqactttcacctqattqttaaaqataaaactq 649
            |||||||||||||||||||||||||||||||||||||||||||||||| ||| ||  ||
Db    49191 CAAAGTCATCTCTGAAAACGGAGAGCTTTGGACTTTCACCTGATTGTTAAAGCTAATGCTC 49132

Qy      650 t 650
            |
Db    49131 T 49131


RESULT    3
AC090685/c
LOCUS       AC090685               171136 bp    DNA     linear   PRI 27-SEP-2001
DEFINITION  Homo sapiens chromosome 17, clone RP11 252O24, complete sequence.
ACCESSION   AC090685
VERSION     AC090685.7  GI:15789269
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171136)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-252O24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171136)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boquslavkiy,L., Boukhqalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodqe,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galaqan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Haqos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenqa,V.,
            Murphy,T., Naylor,J., Nquyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Roqov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Souqnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171136)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boquslavkiy,L., Boukhqalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chanq,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodqe,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galaqan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Haqos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenqa,V., Murphy,T., Naylor,J., Nquyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Roqov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 27, 2001 this sequence version replaced gi:15421875.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10758
                Center clone name: 252_O_24
            --------------.
FEATURES             Location/Qualifiers
     source          1. .171136
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-252O24"
                     /clone_lib="RPCI 11 Human Male BAC"
     repeat_region   496. .810
                     /rpt_family="AluSx"
     repeat_region   complement(903. .1220)
                     /rpt_family="AluSx"
     repeat_region   1958. .2259
                     /rpt_family="AluSx"
     repeat_region   3521. .3541
                     /rpt_family="AT_rich"
     repeat_region   complement(4906. .5206)
                     /rpt_family="AluSx"
     repeat_region   6440. .6732
                     /rpt_family="AluJb"
     repeat_region   6737. .6776
                     /rpt_family="MER4E"
     repeat_region   6794. .6971
                     /rpt_family="AluJb"
     repeat_region   7119. .7242
                     /rpt_family="AluSq"
     repeat_region   7243. .7273
                     /rpt_family="(CAAAA)n"
     repeat_region   7274. .7440
                     /rpt_family="AluSq"
     repeat_region   complement(7666. .7865)
                     /rpt_family="MIR"
     repeat_region   complement(7983. .8112)
                     /rpt_family="MIR"
```

```
repeat_region   complement(8499. .8785)
                /rpt_family="L1ME3"
repeat_region   9605. .9645
                /rpt_family="AT_rich"
repeat_region   10172. .10398
                /rpt_family="L2"
repeat_region   complement(10430. .10704)
                /rpt_family="AluSx"
repeat_region   11396. .11560
                /rpt_family="AluJb"
repeat_region   complement(12217. .12517)
                /rpt_family="AluJo"
repeat_region   12807. .12849
                /rpt_family="(T)n"
repeat_region   13056. .13161
                /rpt_family="MER5A"
repeat_region   13340. .13409
                /rpt_family="L2"
repeat_region   complement(13413. .13530)
                /rpt_family="FRAM"
repeat_region   complement(13613. .13679)
                /rpt_family="MER5B"
repeat_region   complement(13811. .13933)
                /rpt_family="FLAM_A"
repeat_region   complement(13960. .14253)
                /rpt_family="AluSx"
repeat_region   14620. .14901
                /rpt_family="AluJo"
repeat_region   complement(15259. .15559)
                /rpt_family="AluSc"
repeat_region   complement(15645. .15871)
                /rpt_family="MIR"
repeat_region   complement(15948. .16005)
                /rpt_family="L2"
repeat_region   complement(16005. .16129)
                /rpt_family="MIR"
repeat_region   16580. .16864
                /rpt_family="L2"
repeat_region   18380. .18407
                /rpt_family="AT_rich"
repeat_region   complement(18780. .19066)
                /rpt_family="AluJo"
repeat_region   19130. .19163
                /rpt_family="U2"
repeat_region   complement(19812. .19917)
                /rpt_family="U6"
repeat_region   complement(20095. .20376)
                /rpt_family="AluJb"
repeat_region   20423 .20727
                /rpt_family="AluSx"
repeat_region   20848. .20872
                /rpt_family="AT_rich"
repeat_region   23695. .23721
                /rpt_family="(CGG)n"
repeat_region   complement(25012. .25643)
                /rpt_family="L2"
repeat_region   complement(25677. .25939)
                /rpt_family="L2"
repeat_region   25998. .26149
                /rpt_family="MER91A"
repeat_region   complement(26150. .26443)
                /rpt_family="AluSg"
repeat_region   26444. .26493
                /rpt_family="MER91A"
repeat_region   complement(26502. .26750)
                /rpt_family="L2"
repeat_region   27314. .27582
                /rpt_family="AluJo"
repeat_region   complement(27786. .28088)
                /rpt_family="AluSq"
repeat_region   28275. .28362
                /rpt_family="Charlie8"
repeat_region   28363. .28641
                /rpt_family="AluJb"
repeat_region   28642. .28715
                /rpt_family="Charlie8"
repeat_region   28729. .28827
                /rpt_family="AluJb"
repeat_region   28828. .29867
                /rpt_family="LTR28"
repeat_region   29868. .30069
                /rpt_family="AluSq"
repeat_region   complement(30661. .30743)
                /rpt_family="L2"
repeat_region   complement(31351. .31487)
                /rpt_family="L2"
repeat_region   31524. .31656
                /rpt_family="AluSq/x"
repeat_region   complement(31657. .31951)
```

```
  Query Match             36.3%;  Score 286.6;  DB 9;  Length 171136;
  Best Local Similarity   97.0%;  Pred. No. 4.8e-73;
  Matches  292;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     350 tgaatgttacttttcctcctctcctaagtggaaacgacttccaaacaqttgaqgaaqgca 409
           ||| | || |  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56860 TGAGTCTTGCAGTTCCTCCTCTCCTAAGTGGAAACGACTTCCAAACAGTTGAGGAAGGCA 56801

Qy     410 gtaatgtgaaqttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtaca 469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56800 GTAATGTGAAGTTGGTTTGCAATGTGAAAGCCAACCCCCAGGCTCAAATGATGTGGTACA 56741

Qy     470 aaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagt 529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56740 AAAACAGTAGTCTCCTCGATTTAGAGAAAAGCCGTCACCAAATCCAACAGACAAGTGAGT 56681

Qy     530 cttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattg 589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56680 CTTTTCAGCTGTCAATCACCAAAGTCGAGAAGCCTGACAACGGAACCTACAGTTGTATTG 56621

Qy     590 caaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagataaaactg 649
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||  ||
Db   56620 CAAAGTCATCTCTGAAAACGGAGAGCTTGGACTTTCACCTGATTGTTAAAGGTAATGCTC 56561

Qy     650 t 650
           |
Db   56560 T 56560
```

```
RESULT   4
AC087644/c
LOCUS       AC087644               172045 bp    DNA     linear   HTG 08-JAN-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-403E9 map 17, *** SEQUENCING
            IN PROGRESS ***, 8 unordered pieces.
ACCESSION   AC087644
VERSION     AC087644.3  GI:18072182
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172046)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-403E9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172046)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
```

```
            McPherson,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norton,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,K., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Triqilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,R., Wu,X., Wyman,D., Ye,W.J., Yoakum,S., Zaitszun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 6, 2002 this sequence version replaced gi:13357327.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11984
                Center clone name: 403_E_9
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    47343: contig of 47343 bp in length
            *    47344 47443: gap of      100 bp
            *    47444    51984: contig of 4541 bp in length
            *    51985 52084: gap of      100 bp
            *    52085    56414: contig of 4330 bp in length
            *    56415 56514: gap of      100 bp
            *    56515    61504: contig of 5090 bp in length
            *    61505 61604: gap of      100 bp
            *    61605    64448: contig of 2844 bp in length
            *    64449 64548: gap of      100 bp
            *    64549    66982: contig of 2434 bp in length
            *    66983 67082: gap of      100 bp
            *    67083    69497: contig of 2415 bp in length
            *    69498 69597: gap of      100 bp
            *    69598   172046: contig of 102449 bp in length.
FEATURES            Location/Qualifiers
     source         1..172046
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="17"
                    /map="17"
                    /clone="RP11-403E9"
                    /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    50159 a   36414 c   35904 g   48710 t     859 others
ORIGIN


  Query Match             86.4%;  Score 286.6;  DB 2;  Length 172046;
  Best Local Similarity   97.0%;  Pred. No. 4.8e-73;
  Matches  292,  Conservative    0,  Mismatches    9,  Indels    0;  Gaps    0,

Qy     350 tgaatgttactttttctctctctcttaagtggaaacgacttccaaacagttgaggaaggca 409
           ||| | || |  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154978 TGAGTCTTGCAGTTCCTCCTCTCCTAAGTGGAAACGACTTCCAAACAGTTGAGGAAGGCA 154919

Qy     410 gtaatgtgaaqttqqtttqcaatgtgaaaqccaaccccccaqgctcaaatgatgtggtaca 469
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154918 GTAATGTGAAGTTGGTTTGCAATGTGAAAGCCAACCCCCCAGGCTCAAATGATGTGGTACA 154859

Qy     470 aaaacaqtaqtctcctcqatttaqqqaaaaqccqtaccaaatccaacaqacaaqtqaqt 529
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154858 AAAACAGTAGTCTCCTCGATTTAGAGAAAAGCCGTCACCAAATCCAACAGACAAGTGAGT 154799

Qy     530 ctttcaqctqtcaatcaccaaaqtcqaqaaqcctqacaacqqaaccctacaqttqtattq 589
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 154798 CTTTCAGCTGTCAATCACCAAAGTCGAGAAGCCTGACAACGGAACCCTACAGTTGTATTG 154739

Qy     590 caaaqtcatctctqaaaacqqaqaqcttqqacttt.acctqattqttaaaqqtaatqctc 649
           |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db 154738 CAAAGTCATCTCTGAAAACGGAGAGCTTGGACTTTCACCTGATTGTTAAAGGTAATGCTC 154679

Qy     650 t 650
           |
Db 154678 T 154678


RESULT   5
AC024065/c
LOCUS       AC024065              172622 bp    DNA     linear   HTG 17-AUG-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-403E9, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC024065
VERSION     AC024065.5  GI:9838183
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172622)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172622)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 17, 2000 this sequence version replaced gi:9502467.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0403E09
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 100%
            Sequencing vector, plasmid; 0%
            Chemistry: Dye primer ET; 100% of reads
            Chemistry: Dye terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 167048 bases at least Q40
            Consensus quality, 169424 bases at least Q30
            Consensus quality, 170452 bases at least Q20
            Insert size: 196000; agarose-fp
            Insert size: 173251; sum-of-contigs
            Quality coverage, 5.19 in Q20 bases; agarose-fp
            Quality coverage: 5.88 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6930: contig of 6930 bp in length
            *     6931     7030: gap of unknown length
            *     7031    24747: contig of 17717 bp in length
```

```
            *    24748    24847: gap of unknown length
            *    24848    35273: contig of 10426 bp in length
            *    35274    35373: gap of unknown length
            *    35374    56929: contig of 21556 bp in length
            *    56930    57029: gap of unknown length
            *    57030    73825: contig of 16796 bp in length
            *    73826    73925: gap of unknown length
            *    73926    89138: contig of 15213 bp in length
            *    89139    89238: gap of unknown length
            *    89239   117743: contig of 28505 bp in length
            *   117744   117843: gap of unknown length
            *   117844   172622: contig of 54779 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172622
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-403E9"
     misc_feature    1. .6930
                     /note="assembly_name:Contig10"
     misc_feature    7031. .24747
                     /note="assembly_name:Contig11"
     misc_feature    24848. .35273
                     /note="assembly_name:Contig12"
     misc_feature    35374. .56929
                     /note="assembly_name:Contig13
                     clone_end:SP6
                     vector_side:right"
     misc_feature    57030. .73825
                     /note="assembly_name:Contig14"
     misc_feature    73926. .89138
                     /note="assembly_name:Contig15"
     misc_feature    89239. .117743
                     /note="assembly_name:Contig16"
     misc_feature    117844. .172622
                     /note="assembly_name:Contig17
                     clone_end:T7
                     vector_side:right"
BASE COUNT    49385 a  36666 c  36708 g  49156 t    707 others
ORIGIN


  Query Match             36.0%;  Score 284;  DB 2;  Length 172622;
  Best Local Similarity   91.2%;  Pred. No. 2.8e-72;
  Matches  299;  Conservative    2;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      36 aagatttcttctcttagtaatttttatttctgccacgtgagatgacaagttctgttttaac 95
           | | | | | ||||   ||  |  ||   | |  | |    |  || ||||||||||||||
Db 160044 AGGTTGTTTACTCTGTATATCTAAATAGTTTCTCCCTATACTCCCAGGTTCTGTTTTAAC 159985

Qy      96 tgtgaatggtaaaactgagaactatatcctggatactacacctggctcccaagcatctct 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159984 TGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCTGGCTCCCAAGCATCTCT 159925

Qy     156 gatatgtgctgttcaaaaccacaccagagaggaagaactgctctggtaccgagaggaggg 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159924 GATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTCTGGTACCGAGAGGAGGG 159865

Qy     216 gagagtggatttgaaatctgsaaacaamatcaattccagctctgtctgtgtctcttccat 275
           ||||||||||||||||||:||||||:||||||||||||||||||||||||||||||||||
Db 159864 GAGAGTGGATTTGAAATCTGGAAACAAAATCAATTCCAGCTCTGTCTGTGTCTCTTCCAT 159805

Qy     276 cagtgaaaatgacaacggaatcagctttacctgcaggctggggagggatcagtccgtgtc 335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 159804 CAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGGAGGGATCAGTCCGTGTC 159745

Qy     336 cgtttcggtggtgctgaatgttacttttt 363
           ||||||||||||||||||||||||||| |
Db 159744 CGTTTCGGTGGTGCTGAATGTTACTTGT 159717


RESULT   6
AX245713
LOCUS       AX245713                 338 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 643 from Patent WO0166753.
ACCESSION   AX245713
VERSION     AX245713.1  GI:15860387
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 643 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       93 a     69 c     91 g     85 t
ORIGIN


  Query Match             35.2%;  Score 277.4;  DB 6;  Length 338;
  Best Local Similarity   98.2%;  Pred. No. 1.8e-70;
  Matches  278;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      79 acaagttctgttttaactgtgaatggtaaaactgagaactatatcctggatactacacct 138
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 AGAGGTTCTGTTTTAACTGTGAATGGTAAAACTGAGAACTATATCCTGGATACTACACCT 115

Qy     139 ggctcccaagcatctctgatatgtgctgttcaaaaccacaccagagaggaagaactgctc 198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 GGCTCCCAAGCATCTCTGATATGTGCTGTTCAAAACCACACCAGAGAGGAAGAACTGCTC 175

Qy     199 tggtaccgagaggagggggagagtggatttgaaatctgsaaacaamatcaattccagctct 258
           ||||||||||||||||||||||||||||||||||||||:|||||:||||||||||||||||
Db     176 TGGTACCGAGAGGAGGGGGAGAGTGGATTTGAAATCTGGAAACAAAATCAATTCCAGCTCT 235

Qy     259 gtctgtgtctcttccatcagtgaaaatgacaacggaatcagctttacctgcaggctgggg 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 GTCTGTGTCTCTTCCATCAGTGAAAATGACAACGGAATCAGCTTTACCTGCAGGCTGGGG 295

Qy     319 agggatcagtccgtgtccgtttcggtggtgctgaatgttactt 361
           |||||||||||||||||||||||| ||||||||||||||||||
Db     296 AGGGATCAGTCCGTGTCCGTTTCGGAGGTGCTGAATGTTACTT 338


RESULT   7
AX246249
LOCUS       AX246249                 361 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 1179 from Patent WO0166753.
ACCESSION   AX246249
VERSION     AX246249.1  GI:15860923
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 1179 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .361
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       114 a      63 c      90 g      94 t
ORIGIN


  Query Match             31.8%;  Score 250.8;  DB 6;  Length 361;
  Best Local Similarity   98.8%;  Pred. No. 1.2e-62;
  Matches  249;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggcatggaagagcagtgtcataatgcaaatgggaagatttcttctcttagtaatttta 60
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db     109 ATGGCATGGAAGAGCAGTGTCATAATGCAAATGGGAAGATTTCTTCTCTTAGTAATTTTA 168

Qy      61 tttctgccacgtgagatgacaaagttctgttttaactgtgaatggtaaaactgagaactat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     169 TTTCTGCCACGTGAGATGACAAGTTCTGTTTTAACTGTGAATGGTAAAACTGAGAACTAT 228

Qy     121 atcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacacc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     229 ATCCTGGATACTACACCTGGCTCCCAAGCATCTCTGATATGTGCTGTTCAAAACCACACC 288

Qy     181 agagaggaagaactgctctggtaccgagaggagggggagagtggatttgaaatctggaaac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db     289 AGAGAGGAAGAACTGCTCTGGTACCGAGAGGAGGGGAGAGTGGATTTGAAATCTGGAAAC 348

Qy     241 aaatcaattccaac 255
           ||:|||  |||||
Db     349 AAAATCAATTCC 360


RESULT   8
AX246408
LOCUS       AX246408                328 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 1338 from Patent WO0166753.
ACCESSION   AX246408
VERSION     AX246408.1  GI:15861082
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 1338 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       102 a      62 c      81 g      83 t
ORIGIN


  Query Match             31.6%;  Score 249;  DB 6;  Length 328;
  Best Local Similarity   97.6%;  Pred. No. 4e-62;
  Matches  249;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 atggcatggaagagcagtgtcataatgcaaatgggaagatttcttctcttagtaatttta 60
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db      74 ATGGCATGGAAGAGCAGTGTCATAATGCAAATGGGAAGATTTCTTCTCTTAGTAATTTTA 132

Qy      61 tttctgccacgtgagatgacaaagttctgttttaactgtgaatggtaaaactgagaactat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 TTTCTGCCACGTGAGATGACAAGTTCTGTTTTAACTGTGAATGGTAAAACTGAGAACTAT 192

Qy     121 atcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacacc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
             .
```

```
Db     193 ATCCTGGATACTACACCTGGCTCCCAAGCATCTCTGATATGTGCTGTTCAAAACCACACC 252

Qy     181 agagaggaagaactgctctggtaccgagaggagggggagagtggatttgaaatctggaaac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db     253 AGAGAGGAAGAACTGCTCTGGTACCGAGAGGAGGGGAGAGTGGATTTGAAATCTGGAAAC 312

Qy     241 aaatcaattccaac 255
           ||:|||  |||||
Db     313 AAAATCCATTCCCGC 327


RESULT   9
AX070657
LOCUS       AX070657                359 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 1129 from Patent WO0102568.
ACCESSION   AX070657
VERSION     AX070657.1  GI:12580442
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
            Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Lamson,G., Drmanac,R., Crkenjakov,R., Drmanac,S., Dickson,M.,
            Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and Strache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0102568-A 1129 11-JAN-2001;
            CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .359
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    1. .359
                     /note="n = A,T,C or G"
BASE COUNT       113 a      64 c      92 g      89 t       1 others
ORIGIN


  Query Match             31.3%;  Score 247.2;  DB 6;  Length 359;
  Best Local Similarity   98.4%;  Pred. No. 1.4e-61;
  Matches  246;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 atggcatggaagagcagtgtcataatgcaaatgggaagatttcttctcttagtaatttta 60
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db     109 ATGGCATGGAAGAGCAGTGTCATAATGCAAATGGGAAGATTTCTTCTCTTAGTAATTTTA 168

Qy      61 tttctgccacgtgagatgacaaagttctgttttaactgtgaatggtaaaactgagaactat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     169 TTTCTGCCACGTGAGATGACAAGTTCTGTTTTAACTGTGAATGGTAAAACTGAGAACTAT 228

Qy     121 atcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacacc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     229 ATCCTGGATACTACACCTGGCTCCCAAGCATCTCTGATATGTGCTGTTCAAAACCACACC 288

Qy     181 agagaggaagaactgctctggtaccgagaggagggggagagtggatttgaaatctggaaac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db     289 AGAGAGGAAGAACTGCTCTGGTACCGAGAGGAGGGGAGAGTGGATTTGAAATCTGGAAAC 348

Qy     241 aaatcaatt 250
           ||:|||||
Db     349 AAAATCAATT 358


RESULT  10
AL645479
LOCUS       AL645479              201797 bp    DNA     linear   HTG 01 FEB 2002
DEFINITION  Mus musculus chromosome 11 clone RP23-39603, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL645479
VERSION     AL645479.7  GI:18491374
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 4, 2002 this sequence version replaced gi:18476855.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM39603
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 196898 bases at least Q40
            Consensus quality: 198094 bases at least Q30
            Consensus quality: 198893 bases at least Q20
            Insert size: 200297; sum-of-contigs
            Insert size: 187260; 5.3% error; agarose-fp
            Quality coverage: 7.48x in Q20 bases; sum-of-contigs Quality
            coverage: 8.00x in Q20 bases; agarose-fp
            ---------------.
            * NOTE: This is a 'working draft' sequence
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .201797
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-39603"
                     /clone_lib="RPCI-23"
     misc_feature    1. .11113
                     /note="assembly_fragment:02332
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    11214. .48611
                     /note="assembly_fragment:02747
                     fragment_chain:1"
     misc_feature    48712. .59723
                     /note="assembly_fragment:00080
                     fragment_chain:2"
     misc_feature    59824. .80048
                     /note="assembly_fragment:02274
                     fragment_chain:2"
     misc_feature    80149. .83718
                     /note="assembly_fragment:01802
                     fragment_chain:2"
     misc_feature    83819. .109110
                     /note="assembly_fragment:00974
                     fragment_chain:2"
     misc_feature    109211. .128137
                     /note="assembly_fragment:02689
                     fragment_chain:2"
     misc_feature    128238. .137020
                     /note="assembly_fragment:02530
                     fragment_chain:3"
     misc_feature    137121. .142890
                     /note="assembly_fragment:03193
                     fragment_chain:3"
     misc_feature    142991. .152615
                     /note="assembly_fragment:00674
                     fragment_chain:4"
     misc_feature    152716. .159209
                     /note="assembly_fragment:00421
                     fragment_chain:4"
     misc_feature    159310. .162315
                     /note="assembly_fragment:01093
                     fragment_chain:4"
     misc_feature    162416. .181977
                     /note="assembly_fragment:02178
                     fragment_chain:4"
     misc_feature    182078. .193642
                     /note="assembly_fragment:01392
                     fragment_chain:4"
     misc_feature    193743. .198053
                     /note="assembly_fragment:04059
                     fragment_chain:4"
     misc_feature    198154. .201797
                     /note="assembly_fragment:00302
                     fragment_chain:4
                     clone_end:T7
                     vector_side:right"
BASE COUNT    57371 a   42459 c   42767 g   57700 t    1500 others
ORIGIN


  Query Match             27.2%;  Score 214.6;  DB 2;  Length 201797;
  Best Local Similarity   82.1%;  Pred. No. 7.1e-52;
  Matches  247;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     350 tgaatgttacttttcctcctctcctaagtggaaacgacttccaaacagttgaggaaggca 409
           ||| | ||||  |||||||||| |||||||| || | ||||||||| || ||||||  || 
Db  123219 TGAGTCTTACAGTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACA 123278

Qy     410 gtaatgtgaagttggtttgcaatgtgaaagccaaccccccaggctcaaatgatgtggtaca 469
           |  |||| |  |||||||||||| |||||| ||||||| || || ||||||||||||||| |
Db  123279 GCGATGIAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATGTGGTATA 123338

Qy     470 aaaacagtagtctcctcgatttagagaaaagccgtcaccaaatccaacagacaagtgagt 529
           ||||||  ||   | | | |||||||||| ||||||||||||||| |||||||| ||||
Db  123339 AAAACAACAGCGCCTTGCTTTTAGAGAAAGCCGTCACCAAATCCACCAGACAAGGGAGT 123398

Qy     530 cttttcagctgtcaatcaccaaagtcgagaagcctgacaacggaacctacagttgtattg 589
           |||||||| |||| |||||||||||| ||||  |||||| || |||||||| |||||||
Db  123399 CTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGCTGTATTG 123458

Qy     590 caaagtcatctctgaaaacggagagcttggactttcacctgattgttaaagataaaactg 649
           | |  |||||||||||||| ||||| | ||||||| |||||| ||||||||| |||  ||
Db  123459 CCAGCTCATCTCTGAAAATGGAGACCATGGACTTCCACCTGCTTGTTAAAGGTAACTCTC 123518

Qy     650 t 650
           |
Db  123519 T 123519


RESULT  11
AL603842/c
LOCUS       AL603842               229289 bp    DNA     linear   ROD 17-NOV-2001
DEFINITION  Mouse DNA sequence from clone RP23-100P23 on chromosome 11,
            complete sequence.
ACCESSION   AL603842
VERSION     AL603842.9  GI:16519591
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 229289)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 30, 2001 this sequence version replaced gi:16304764.
            During sequence assembly data is compared from overlapping clones.
```

```
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted, all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep WP:CE19822 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP23-100P23.
FEATURES             Location/Qualifiers
     source          1..229289
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-100P23"
                     /clone_lib="RPCI-23"
     misc_feature    59782..59788
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    125096..125397
                     /note="Single clone region. Sequence from clone PCR only."
     misc_feature    170770..170774
                     /note="1327 bases of IS2 transposon (V00610) removed here.
                     This sequence represents the duplicated flanking sequence
                     of the IS2."
     misc_feature    176557..176641
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    64835 a  50446 c  50226 g  63782 t
ORIGIN


  Query Match             27.2%;  Score 214.6;  DB 10;  Length 229289;
  Best Local Similarity   82.1%;  Pred. No. 7.2e-52;
  Matches  247;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      350 tgaatgttactttttcttctctcctaaqtggaaacqacttccaaaacagttgaqqaaggca 409
            ||| | ||||  ||||||||||| ||||||| || | ||||||||| || ||||||  ||
Db   191296 TGAGTCTTACAGTTCCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACA 191237

Qy      410 gtaatgtgaaqttggtttgcaatgtgaaagccaaccccccaqqctccaaatgatgtggtaca 469
            |  |||| |  |||||||||| |||||| ||||||| || || ||||||||||||||| |
Db   191236 GCGATGTAAGTTTGGTTTGCAACGTGAAATCCAACCCTCAAGCGCAAATGATGTGGTATA 191177

Qy      470 aaaacaqtagtctcctcqatttaqaqaaaaqccqtcaccaaatccaacaqacaaqtgaqt 529
            ||||||  ||    | | | |||||||||| |||||||||||||||| |||||||| ||||
Db   191176 AAAACAACAGCGCCTTGGTTTTAGAGAAAGGCCGTCACCAAATCCACCAGACAAGGGAGT 191117

Qy      530 cttttcaqctgtcaatcaccaaaqtcqaqaaqcctqacaacgqaacctacaqttgtattg 589
            ||||||| |||| |||||||||||| ||||  |||||| || |||||||| |||||||
Db   191116 CTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACAATGGGACCTACAGCTGTATTG 191057

Qy      590 caaaqtcatctctqaaaacqqaqagattggactttaacctgattgttaaaqataaaactg 649
            | |  |||||||||||| |||| | ||||||| |||||| ||||||||| |||  ||
Db   191056 CCAGCTCATCTCTGAAAATGGAGACCATGGACTTGCACCTGCTTGTTAAAGGTAACTCTC 190997

Qy      650 t 650
            |
Db   190996 T 190996
```

```
RESULT  12
AX070653
LOCUS       AX070653                359 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 1125 from Patent WO0102568.
ACCESSION   AX070653
VERSION     AX070653.1  GI:12580438
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia, Eutheria, Primates, Catarrhini, Hominidae; Homo.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
            Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Lamson,G., Drmanac,R., Crkenjakov,R., Drmanac,S., Dickson,M.,
            Labat,I., Teshkowitz,D., Kita,D., Garcia,V. and Strache Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0102568-A 1125 11-JAN-2001;
            CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1..359
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    1..359
                     /note="n = A,T,C or G"
BASE COUNT      106 a     69 c     95 g     88 t      1 others
ORIGIN


  Query Match             25.8%;  Score 203.4;  DB 6;  Length 359;
  Best Local Similarity   94.6%;  Pred. No. 1e-48;
  Matches  229;  Conservative    2;  Mismatches    9;  Indels    2;  Gaps    2;

Qy        1 atggcatggaagagcagtgtcataatgcaratgggaagatttcttctcttagtaatttta 60
            ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||| ||
Db      109 ATGGCATGGAAGAGCAGTGTCATAATGCAAATGGGAAGATTTCTTCTCTTAGTAATTCTA 168

Qy       61 tttctgccacgtgagatgacaagtttctgttttaactgtgaatggtaaaactgagaactat 120
            |||||||||||||||||||||||||||| |||||||||||| |||| ||||||||||||
Db      169 TTTCTGCCACGTGAGATGACAAGTTCTGGTTTAACTGTGAATCGTAACACTGAGAACTAT 228

Qy      121 atcctggatactacacctggctcccaagcatctctgatatgtgctgttcaaaaccacacc 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      229 ATCCTGGATACTACACCTGGCTCCCAAGCATCTCTGATATGTGCTGTTCCAAACCACACC 288

Qy      181 agaggaggaaqaactgctctggtaccgagaggagggaqatggaatttgaaatctgsaaac 240
            |||||||||| |||||||||||||||||||| ||||||| ||||||| |||||:||
Db      289 AGAGAGGAAG-ACTGCTCTGGTACCGAGAGGA-GGGAGAGGGGATTTGANATCTGGAACA 346

Qy      241 aa 242
            ||
Db      347 AA 348


RESULT  13
AC103795
LOCUS       AC103795                70248 bp    DNA     linear   HTG 29-NOV-2001
DEFINITION  Homo sapiens chromosome 11 clone CTD-2529D2 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC103795
VERSION     AC103795.1  GI:17149680
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70248)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone CTD-2529D2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70248)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
```

Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L21770
Center clone name: 2529_D_2
----------.
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 687: contig of 687 bp in length
* 688 787: gap of 100 bp
* 788 1501: contig of 714 bp in length
* 1502 1601: gap of 100 bp
* 1602 2345: contig of 744 bp in length
* 2346 2445: gap of 100 bp
* 2446 3151: contig of 706 bp in length
* 3152 3251: gap of 100 bp
* 3252 3929: contig of 678 bp in length
* 3930 4029: gap of 100 bp
* 4030 4719: contig of 690 bp in length
* 4720 4819: gap of 100 bp
* 4820 5538: contig of 719 bp in length
* 5539 5638: gap of 100 bp
* 5639 6356: contig of 718 bp in length
* 6357 6456: gap of 100 bp
* 6457 7156: contig of 700 bp in length
* 7157 7256: gap of 100 bp
* 7257 7888: contig of 632 bp in length
* 7889 7988: gap of 100 bp
* 7989 8743: contig of 755 bp in length
* 8744 8843: gap of 100 bp
* 8844 9533: contig of 690 bp in length
* 9534 9633: gap of 100 bp
* 9634 10371: contig of 738 bp in length
* 10372 10471: gap of 100 bp
* 10472 11203: contig of 732 bp in length
* 11204 11303: gap of 100 bp
* 11304 12016: contig of 713 bp in length
* 12017 12116: gap of 100 bp
* 12117 12823: contig of 707 bp in length
* 12824 12923: gap of 100 bp
* 12924 13628: contig of 705 bp in length
* 13629 13728: gap of 100 bp
* 13729 14441: contig of 713 bp in length
* 14442 14541: gap of 100 bp
* 14542 15272: contig of 731 bp in length
* 15273 15372: gap of 100 bp
* 15373 16106: contig of 734 bp in length
* 16107 16206: gap of 100 bp
* 16207 16935: contig of 729 bp in length
* 16936 17035: gap of 100 bp
* 17036 17762: contig of 727 bp in length
* 17763 17862: gap of 100 bp
* 17863 18558: contig of 696 bp in length
* 18559 18658: gap of 100 bp
* 18659 19351: contig of 693 bp in length
* 19352 19451: gap of 100 bp
* 19452 20181: contig of 730 bp in length
* 20182 20281: gap of 100 bp
* 20282 21023: contig of 742 bp in length
* 21024 21123: gap of 100 bp
* 21124 21827: contig of 704 bp in length
* 21828 21927: gap of 100 bp
* 21928 22624: contig of 697 bp in length
* 22625 22724: gap of 100 bp
* 22725 23401: contig of 677 bp in length
* 23402 23501: gap of 100 bp
* 23502 24228: contig of 727 bp in length
* 24229 24328: gap of 100 bp
* 24329 25055: contig of 727 bp in length
* 25056 25155: gap of 100 bp
* 25156 25898: contig of 743 bp in length
* 25899 25998: gap of 100 bp
* 25999 26710: contig of 712 bp in length
* 26711 26810: gap of 100 bp
* 26811 27551: contig of 741 bp in length
* 27552 27651: gap of 100 bp
* 27652 28394: contig of 743 bp in length
* 28395 28494: gap of 100 bp
* 28495 29230: contig of 736 bp in length
* 29231 29330: gap of 100 bp
* 29331 30037: contig of 707 bp in length
* 30038 30137: gap of 100 bp
* 30138 30854: contig of 717 bp in length
* 30855 30954: gap of 100 bp
* 30955 31680: contig of 726 bp in length
* 31681 31780: gap of 100 bp
* 31781 32520: contig of 740 bp in length
* 32521 32620: gap of 100 bp
* 32621 33354: contig of 734 bp in length
* 33355 33454: gap of 100 bp
* 33455 34191: contig of 737 bp in length
* 34192 34291: gap of 100 bp
* 34292 35034: contig of 743 bp in length
* 35035 35134: gap of 100 bp
* 35135 35880: contig of 746 bp in length
* 35881 35980: gap of 100 bp
* 35981 36663: contig of 683 bp in length
* 36664 36763: gap of 100 bp
* 36764 37477: contig of 714 bp in length
* 37478 37577: gap of 100 bp
* 37578 38321: contig of 744 bp in length
* 38322 38421: gap of 100 bp
* 38422 39110: contig of 689 bp in length
* 39111 39210: gap of 100 bp
* 39211 39892: contig of 682 bp in length
* 39893 39992: gap of 100 bp
* 39993 40678: contig of 686 bp in length

```
              *    40679 40778: gap of       100 bp
              *    40779    41479: contig of 701 bp in length
              *    41480 41579: gap of       100 bp
              *    41580    42310: contig of 731 bp in length
              *    42311 42410: gap of       100 bp
              *    42411    43144: contig of 734 bp in length
              *    43145 43244: gap of       100 bp
              *    43245    43985: contig of 741 bp in length
              *    43986 44085: gap of       100 bp
              *    44086    44777: contig of 692 bp in length
              *    44778 44877: gap of       100 bp
              *    44878    45598: contig of 721 bp in length
              *    45599 45698: gap of       100 bp
              *    45699    46433: contig of 735 bp in length
              *    46434 46533: gap of       100 bp
              *    46534    47267: contig of 734 bp in length
              *    47268 47367: gap of       100 bp
              *    47368    48058: contig of 691 bp in length
              *    48059 48158: gap of       100 bp
              *    48159    48880: contig of 722 bp in length
              *    48881 48980: gap of       100 bp
              *    48981    49717: contig of 737 bp in length
              *    49718 49817: gap of       100 bp
              *    49818    50560: contig of 743 bp in length
              *    50561 50660: gap of       100 bp
              *    50661    51400: contig of 740 bp in length
              *    51401 51500: gap of       100 bp
              *    51501    52227: contig of 727 bp in length
              *    52228 52327: gap of       100 bp
              *    52328    53066: contig of 739 bp in length
              *    53067 53166: gap of       100 bp
              *    53167    53853: contig of 687 bp in length
              *    53854 53953: gap of       100 bp
              *    53954    54675: contig of 722 bp in length
              *    54676 54775: gap of       100 bp
              *    54776    55508: contig of 733 bp in length
              *    55509 55608: gap of       100 bp
              *    55609    56347: contig of 739 bp in length

  Query Match             13.9%;  Score 109.6;  DB 2;  Length 70248;
  Best Local Similarity   92.2%;  Pred. No. 5e-21;
  Matches  115;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     632 ttqttaaaqataaaactqtqqqtqtaccaataqaqcccattattqctgcatgtgttgtga 691
           ||   || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   44491 TTCCTACAGATAAAACTGTGGGTGTACCAATAGAGCCCATTATTGCTGCATGTGTTGTGA 44450

Qy     692 tctttctqacattqtqctttqqactqattgctaqaaqaaaqaaaatgaatqaaqctctqca 751
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| | |  |
Db   44451 TCTTTCTGACATTGTGCTTTGGACTGATTGCTAGAAGAAAGAAAATAATGAAGGTATCTA 44510

Qy     752 tqaa 755
           | |
Db   44511 AGTA 44514


RESULT  14
AC087440/c
LOCUS       AC087440               217730 bp    DNA     linear   HTG 16-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-425G13 map 11, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC087440
VERSION     AC087440.2  GI:13357319
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 217730)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-425G13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 217730)
```

```
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boquslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,J., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaBoogue,K., LamaZares,E., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenqa,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Travers,M., Travis,N., Triqilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2001 this sequence version replaced gi:12039453.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11822
                Center clone name: 425_G_13
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 192528 bases at least Q40
                Consensus quality: 205654 bases at least Q30
                Consensus quality: 210625 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 215130; sum-of-contigs
                Quality coverage: 4.0 in Q20 bases; agarose-fp
                Quality coverage: 3.4 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   182949: contig of 182949 bp in length
            *  182950 183049: gap of       100 bp
            *  183050   183137: contig of 88 bp in length
            *  183138 183237: gap of       100 bp
            *  183238   183640: contig of 403 bp in length
            *  183641 183740: gap of       100 bp
            *  183741   184713: contig of 973 bp in length
            *  184714 184813: gap of       100 bp
            *  184814   185319: contig of 506 bp in length
            *  185320 185419: gap of       100 bp
            *  185420   186040: contig of 621 bp in length
            *  186041 186140: gap of       100 bp
            *  186141   186820: contig of 680 bp in length
            *  186821 186920: gap of       100 bp
            *  186921   187855: contig of 935 bp in length
            *  187856 187955: gap of       100 bp
            *  187956   188593: contig of 638 bp in length
```

```
            *   188594 188693: gap of       100 bp
            *   188694   189289: contig of 596 bp in length
            *   189290 189389: gap of       100 bp
            *   189390   190822: contig of 1433 bp in length
            *   190823 190922: gap of       100 bp
            *   190923   191663: contig of 741 bp in length
            *   191664 191763: gap of       100 bp
            *   191764   192866: contig of 1103 bp in length
            *   192867 192966: gap of       100 bp
            *   192967   194120: contig of 1154 bp in length
            *   194121 194220: gap of       100 bp
            *   194221   194904: contig of 684 bp in length
            *   194905 195004: gap of       100 bp
            *   195005   195907: contig of 903 bp in length
            *   195908 196007: gap of       100 bp
            *   196008   197287: contig of 1280 bp in length
            *   197288 197387: gap of       100 bp
            *   197388   198887: contig of 1500 bp in length
            *   198888 198987: gap of       100 bp
            *   198988   200386: contig of 1399 bp in length
            *   200387 200486: gap of       100 bp
            *   200487   201763: contig of 1277 bp in length
            *   201764 201863: gap of       100 bp
            *   201864   203900: contig of 2037 bp in length
            *   203901 204000: gap of       100 bp
            *   204001   205876: contig of 1876 bp in length
            *   205877 205976: gap of       100 bp
            *   205977   207807: contig of 1831 bp in length
            *   207808 207907: gap of       100 bp
            *   207908   210489: contig of 2582 bp in length
            *   210490 210589: gap of       100 bp
            *   210590   213765: contig of 3176 bp in length
            *   213766 213865: gap of       100 bp
            *   213866   217087: contig of 3222 bp in length
            *   217088 217187: gap of       100 bp
            *   217188   217730: contig of 543 bp in length.
FEATURES             Location/Qualifiers
     source          1. .217730
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-425G13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .182949
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    183050. .183137
                     /note="assembly_fragment"
     misc_feature    183238. .183640
                     /note="assembly_fragment"
     misc_feature    183741. .184713
                     /note="assembly_fragment"
     misc_feature    184814. .185319
                     /note="assembly_fragment"
     misc_feature    185420. .186040
                     /note="assembly_fragment"
     misc_feature    186141. .186820
                     /note="assembly_fragment"
     misc_feature    186921. .187855
                     /note="assembly_fragment"
     misc_feature    187956. .188593
                     /note="assembly_fragment"
     misc_feature    188694. .189289
                     /note="assembly_fragment"
     misc_feature    189390. .190822
                     /note="assembly_fragment"
     misc_feature    190923. .191663
                     /note="assembly_fragment"
     misc_feature    191764. .192866
                     /note="assembly_fragment"
     misc_feature    192967. .194120
                     /note="assembly_fragment"
     misc_feature    194221. .194904
                     /note="assembly_fragment"
     misc_feature    195005. .195907
                     /note="assembly_fragment"
     misc_feature    196008. .197287
                     /note="assembly_fragment"
     misc_feature    197388. .198887
                     /note="assembly_fragment"
     misc_feature    198988. .200386
                     /note="assembly_fragment"
     misc_feature    200487. .201763
                     /note="assembly_fragment"
     misc_feature    201864. .203900
                     /note="assembly_fragment"
     misc_feature    204001. .205876
                     /note="assembly_fragment"
     misc_feature    205977. .207807
                     /note="assembly_fragment"
     misc_feature    207908. .210489
                     /note="assembly_fragment"
     misc_feature    210590. .213765
                     /note="assembly_fragment"
     misc_feature    213866. .217087
                     /note="assembly_fragment"
     misc_feature    217188. .217730
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    69080 a  40450 c  39974 g  65624 t   2602 others
ORIGIN


  Query Match             5.6%;  Score 43.8;  DB 2;  Length 217730;
  Best Local Similarity   95.7%;  Pred. No. 0.12;
  Matches   45;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     743 agctctgcatgaaggataaagaccctcacagtgaaacagctctatga 789
           ||||||||||||||||||||||||||||||||||||||||||||| | |
Db  199852 AGCTCTGCATGAAGGATAAAGACCCTCACAGTGAAACAGCTCTGTAA 199806


RESULT  15
AF429315/c
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
```

```
                     /note="isolated from a patient with Huntington's
                     Disease Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note "JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGREAFDGGSYGGGWEDGKAHGHGVCIGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29956 a   32731 c   30696 g   28283 t    4254 others
ORIGIN
```

Query Match 5.3%; Score 42; DB 9; Length 125020;
Best Local Similarity 8.6%; Pred. No. 0.39;
Matches 36; Conservative 211; Mismatches 170; Indels 0; Gaps 0;

```
Qy      166 qttcaaaaccacaccaqaqaqqaaqaactqctctqqtaccqaqaqqaggggaqaqtggat 225
            :::::|:  ::|:::   |  | ::::  |::  |::::::: :::|  :    : ::
Db    61034 DKDHHAWSKSRCVHSGASMACWATVKVHGWCWVGWGLDMMSWSCKKMGWMKCMSHDWRDR 60975

Qy      226 ttqaaatctqsaaacaamatcaattccaqctctqtctqtqtctcttccatcagtqaaaat 285
            ::::::  :|: :: : : ::::::    :: : ::::: :::::::: :::: : :  ::
Db    60974 HHDVWMCAKGRCWDGMBWSYSWMWDGWBHMKWDHBBHDBWYSWVWDBDRDBWMWMWCBRD 60915

Qy      286 qacaacqqaatcaqctttacctqcaqqctqqqqaqqqatcaqtccqtqtccqtttcggtq 345
            :: : ::  :::  :: :     ::  :::::  :  :: :: :::   ::    : :
Db    60914 RDDHYYDYSDBBBWHYVYYWKNHBWYHHWSSWYDYHSWSHHNYHVWRMHVRMVBVHHNMN 60855

Qy      346 qtqctqaatqttacttttcctcctctcctaaqtqqaaacqacttccaaacaqttqaqqaa 405
             :: ::::    ::   : : ::  :::::: : : ::: : ::   ||:: :|    :
Db    60854 MWVDWSMMMMVNMHVRVBABMVMMNBSMWVVMWNDNMRYWMAHKRGCAAMMMKTMSACMT 60795

Qy      406 qqcaqtaatqtqaaqttqqtttqcaatqtqaaaqccaaccccccaqqctcaaatqatqtqq 465
              : : ::|    :: : :    ::  |: ::|::| :  ::::   :|: |: : ::
Db    60794 TMSCRSRWTCAAGWRAKWMKMMYRCRWMAGMCWSCMRARYWKMWRRGAWWAMNGRRKRSS 60735

Qy      466 tacaaaaacaqtaqtctcctcqatttaqaqaaaaqccqtcaccaaatccaaacagacaagt 525
             :| ::::: : : ::   : : :   : :: :|::     :: :::: :| |: : :::
Db    60734 CMCCRMMRMAMWGRRMWWAWWRYWRMMWWKYWARKAGARAMSRRRWYTMACCMWMWWRRC 60675

Qy      526 qaqtcttttcaqctqtcaatcaccaaaqtcqaqaaqcctqacaacggaacctacaqt 582
             :: |:  :|::::      :|:|::|: ::|   :|:: :: ::|:  || | |||
Db    60674 MRRRCYGRWCWSSWACWYCMSAMCWMARRMSAWYYRCSYAMMYMMGSCTCCCAAAGT 60618
```

Search completed: September 17, 2002, 10:04:44
Job time: 6192 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 10:01:13 ; Search time 86.46 Seconds
(without alignments)
326.310 Million cell updates/sec

Title: US-09-863-823-2
Perfect score: 1295
Sequence: 1 MQMGPFLLLVILFLPREMTS.....RKKIMKLCMKDKDPHSETAL 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	311	24.0	66	22	AAU18000	Human immunoglobul
2	171	13.2	1571	19	AAW42087	Human Down syndrom
3	170	13.1	1910	19	AAW42086	Human Down syndrom
4	169.5	13.1	359	22	ABB63920	Drosophila melanog
5	169.5	13.1	431	20	AAW94996	SAF-2 polypeptide
6	169.5	13.1	431	20	AAW94995	SAF-2 polypeptide.
7	166	12.8	848	21	AAY88565	Human NCAM 140kD i
8	162	12.5	404	22	AAB61142	Human NOV12 protei
9	161	12.4	1079	22	ABB71337	Drosophila melanog
10	158	12.2	444	20	AAY33741	Beta-secretase H
11	158	12.2	444	22	AAB47251	Beta-secretase H
12	157.5	12.2	1415	22	AAB50692	C. elegans UNC-40
13	157	12.1	467	22	ABB63044	Drosophila melanog
14	155	12.0	595	22	AAB70070	Human secreted pro
15	154.5	11.9	537	22	AAG77857	Mutant protein mMu
16	154.5	11.9	576	22	AAG77858	Mutant protein mMu
17	154.5	11.9	869	22	AAG77856	Protein of muscle
18	154.5	11.9	869	22	AAB68421	Amino acid sequenc
19	154	11.9	582	17	AAR92256	Neural cell adhesi
20	154	11.9	761	17	AAR92255	Neural cell adhesi
21	154	11.9	868	18	AAW26610	Rat muscle-specifi
22	154	11.9	868	18	AAW26507	Rat Dmk receptor.
23	153	11.8	467	19	AAW59992	Sialoadhesin famil
24	153	11.8	467	21	AAY97218	Sialic acid bindin
25	153	11.8	467	21	AAB25581	CD33-like 3 protei
26	153	11.8	467	22	AAM23894	Human EST encoded
27	153	11.8	467	22	AAU14345	Human novel protei
28	153	11.8	467	22	AAU14581	Human novel protei
29	153	11.8	467	22	AAU14582	Human novel protei
30	153	11.8	467	22	AAB29188	Siglec 5 protein.
31	152	11.7	322	22	AAE04367	Human kinase (PKIN
32	152	11.7	869	18	AAW26611	Human muscle-speci
33	152	11.7	869	18	AAW26506	Human Dmk receptor
34	151.5	11.7	983	21	AAY70775	Follistatin-relate
35	149	11.5	478	17	AAR92718	Mouse muscle-local
36	149	11.5	860	17	AAR92716	Mouse muscle-local
37	149	11.5	868	17	AAR92717	Mouse muscle-local
38	149	11.5	1447	16	AAR68553	Deleted in colorec
39	149	11.5	1447	20	AAY33498	Human DCC protein.
40	149	11.5	1447	22	AAB50693	Human UNC-40 prote
41	149	11.5	1728	12	AAR13144	Deleted in Colorec
42	149	11.5	2037	22	ABB71328	Drosophila melanog
43	148.5	11.5	245	19	AAW62577	Alternatively spli
44	148.5	11.5	308	17	AAW05157	Human LAMP residue
45	148.5	11.5	325	17	AAW05152	Human LAMP residue

ALIGNMENTS

RESULT 1
AAU18000
ID AAU18000 standard; Protein; 66 AA.
XX
AC AAU18000;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immunoglobulin polypeptide SEQ ID No 145.
XX
KW Immunoglobulin; signal transduction pathway protein; cancer;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
OS Homo sapiens.
XX
PN WO200155315-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01326.
XX
PP 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229443.
PR 01-SEP-2000; 2000US-0229444.
PR 01-SEP-2000; 2000US-0229445.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0241968.
PR 14-SEP-2000; 2000US-0242397.
PR 14-SEP-2000; 2000US-0242398.
PR 14-SEP-2000; 2000US-0242399.
PR 14-SEP-2000; 2000US-0242400.
PR 14-SEP-2000; 2000US-0242401.
PR 14-SEP-2000; 2000US-0243063.
PR 14-SEP-2000; 2000US-0243064.
PR 14-SEP-2000; 2000US-0243065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234907.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237036.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251088.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0256078.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-157725/49.
DR N-PSDB; AAS28788.
XX
PT Isolated novel immunoglobulin polypeptide for monitoring the presence
PT and progression of diseases and for diagnosis -
XX
PS Claim 11; SEQ ID No 145; 551pp; English.
XX
CC Sequences AAU17977 AAU18087 represent immunoglobulin polypeptides of the
CC invention. The polypeptides and their associated polynucleotides can be
CC used to diagnose a pathological condition or a susceptibility to a
CC pathological condition in a subject by determining the presence or
CC absence of a mutation in a DNA sequence or determining the presence or
CC amount of expression of the protein. Alternatively the identification of
CC a binding partner to a sequence allows determination of changes in

```
CC   protein activity. The sequences can be used as research tools for
CC   receptors or other signal transduction pathway proteins that interact
CC   with the polypeptides of the invention and can be used to treat, prevent
CC   or diagnose various types of disorders such as neurological disorders,
CC   cardiovascular disorders, gastrointestinal disorders, reproductive
CC   disorders, immune system disorders, renal disorders, muscular disorders,
CC   pulmonary disorders, proliferative disorders and cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   66 AA;


  Query Match             24.0%;  Score 311;  DB 22;  Length 66;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-20;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     145 MMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIV 204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mmwyknssildleksrhqiqqtsesfqlsitkvekpdngtysciakssiktesldfhliv 60

Qy     205 K 205
           |
Db      61 k 61


RESULT   2
AAW42087
ID   AAW42087 standard; Protein; 1571 AA.
XX
AC   AAW42087;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Human Down syndrome-cell adhesion molecule DS-CAM2.
XX
KW   DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;
KW   signal transduction; trisomy 21; mental retardation;
KW   holoprosencephaly; corpus callosum agenesis;
KW   schizencephaly; diagnosis; assay; human.
XX
OS   Homo sapiens.
XX
PN   WO9817795-A1.
XX
PD   30-APR-1998.
XX
PF   23-OCT-1997;   97WO-US19547.
XX
PR   25-OCT-1996;   96US-0029322.
XX
PA   (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI   Korenberg JR;
XX
DR   WPI; 1998-271791/24.
DR   N-PSDB; AAV31988.
XX
PT   New isolated Down's Syndrome-cell adhesion molecule - used to
PT   develop products for detection, diagnosis and therapy of
PT   developmental and neurological abnormalities
XX
PS   Claim 2; Page 90-95; 109pp; English.
XX
CC   This polypeptide comprises Down syndrome-cell adhesion molecule
CC   DS-CAM2, an extracellular soluble protein belonging to a novel
CC   subclass of the Ig superfamily with highest homology to neural cell
CC   adhesion molecules.  Its amino acid sequence was deduced from cDNA
CC   clones (see AAV31982) isolated from a trisomy 21 foetal brain library.
CC   It is a splice variant of membrane-bound DS-CAM1 (see AAW42086), and
CC   lacks the entire transmembrane domain of DS-CAM1.  The invention
CC   provides human and murine DS-CAM nucleic acid sequences (see also
CC   AAV31981, AAV31985-87), expression vectors and host cells, transgenic
CC   animals, antibodies, antisense oligonucleotides, and primers
CC   derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated
CC   with developmental and neurological processes.  They can be used in
CC   e.g. neural prosthetic devices used in entubulation methods of
CC   repairing (regenerating) damaged or severed peripheral nerves, and
CC   also in bioassays to identify agonists and antagonists.  The products
CC   can also be used in detection, diagnosis and therapy of developmental
CC   and neurological abnormalities such as Down syndrome, mental
CC   retardation, holoprosencephaly, agenesis of the corpus callosum,
CC   or schizencephaly.
XX
SQ   Sequence   1571 AA;


  Query Match             13.2%,  Score 171,  DB 19,  Length 1571;
  Best Local Similarity   30.4%;  Pred. No. 7.5e-06;
  Matches   58;  Conservative   30;  Mismatches   69;  Indels   34;  Gaps   10;

Qy      15 PREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRVDLKSGNK- 73
           ||:: |||                 ||| || |:|   | ::|| |||  | : |  |
Db     320 prkvkssv---------------gsqvslscsvtg-tedqelswyr-ngei-lnpqknv 360

Qy      74 ----INSSSVCVSSISENDNGISFTCRLGRDQSVS---VSVVLNVTFPPLLSGNDFQTVE 126
               ||  :: :  : ::| | :: | : :|:  :   | |||    | ::|    : |
Db     361 ritqinhenlimdhmvksdgg-ayqcfvrkdklsaqdyvqvvledqtpkiisafsekvvs 419

Qy     127 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQ--TSESFQLS---ITKVEKPD 181
           | |:||||  |   : |  :   : |:   |:| |  |||   ||   |:  :  |
Db     420 paepvslmcnvkgtplptitwtldddpi lkqgshrisqmitseqnvlsylnisssqvrd 478

Qy     182 NGTYSCIAKSS 192
           | | | | :|
Db     479 ggvyrctanns 489


RESULT   3
AAW42086
ID   AAW42086 standard; Protein; 1910 AA.
XX
AC   AAW42086;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Human Down syndrome-cell adhesion molecule DS-CAM1.
XX
KW   DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
KW   signal transduction; trisomy 21; mental retardation;
KW   holoprosencephaly; corpus callosum agenesis;
KW   schizencephaly; diagnosis; assay; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Sig_peptide
FT   Protein         24..1910
FT                   /label= Mat_protein
FT   Domain          24..887
FT                   /label= IG
FT                   /note= "immunoglobulin type-C2 domain"
FT   Domain          888..1594
FT                   /label= FbN
FT                   /note= "fibronectin type III domain"
FT   Domain          1595..1616
FT                   /label= Transmembrane
FT   Domain          1617..1910
FT                   /label= Cytoplasmic
FT   Region          24..126
FT                   /label= Ig1
FT   Region          127..225
FT                   /label= Ig2
FT   Region          226..316
```

```
FT                   /label= Iq3
FT   Region          317..409
FT                   /label= Iq4
FT   Region          410..506
FT                   /label= Iq5
FT   Region          507..603
FT                   /label= Iq6
FT   Region          604..697
FT                   /label= Iq7
FT   Region          698..792
FT                   /label= Iq8
FT   Region          793..887
FT                   /label= Iq9
FT   Disulfide-bond  46..102
FT   Disulfide-bond  145..197
FT   Disulfide-bond  245..293
FT   Disulfide-bond  335..385
FT   Disulfide-bond  428..484
FT   Disulfide-bond  525..575
FT   Disulfide-bond  617..669
FT   Disulfide-bond  711..766
FT   Disulfide-bond  809..865
FT   Disulfide-bond  1307..1359
FT   Modified-site   78..80
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   106..108
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   470..472
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   487..489
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   658..660
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   666..668
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   710..712
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   748..750
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   795..797
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   924..926
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1142..1144
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1160..1162
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1250..1252
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1271..1273
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1324..1326
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1341..1343
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1488..1490
FT                   /note= "Asn is N-glycosylated"
XX
PN   WO9817795-A1.
XX
PD   30-APR-1998.
XX
PF   24-OCT-1997;    97WO-US19547.
XX
PR   25-OCT-1996;    96US-0029422.
XX
PA   (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI   Korenberg JR;
XX
DR   WPI; 1998-271791/24.
DR   N-PSDB; AAV31981.
XX
PT   New isolated Down's Syndrome-cell adhesion molecule - used to
PT   develop products for detection, diagnosis and therapy of
PT   developmental and neurological abnormalities
XX
PS   Claim 2; Page 73-78; 109pp; English.
XX
CC   This polypeptide comprises Down syndrome cell adhesion molecule
CC   DS-CAM1, a cell surface glycoprotein belonging to a novel subclass
CC   of the Ig superfamily with highest homology to neural cell adhesion
CC   molecules.  Its amino acid sequence was deduced from cDNA clones
CC   (see AAV31981) isolated from a trisomy 21 foetal brain library.  A
CC   splice variant, DS-CAM2 (see AAW42087), which is non-membrane bound
CC   was also identified.  The invention also provides human and murine
CC   DS-CAM nucleic acid sequences (see also AAV31985-88), expression
CC   vectors and host cells, transgenic animals, antibodies, antisense
CC   oligonucleotides, and primers derived from DS CAM nucleic acid.
CC   DS CAM polypeptides are associated with developmental and
CC   neurological processes.  They can be used in e.g. neural prosthetic
CC   devices used in entubulation methods of repairing (regenerating)
CC   damaged or severed peripheral nerves, and also in bioassays to
CC   identify agonists and antagonists.  The products can also be
CC   used in detection, diagnosis and therapy of developmental and
CC   neurological abnormalities such as Down syndrome, mental
CC   retardation, holoprosencephaly, agenesis of the corpus callosum,
CC   or schizencephaly.
XX
SQ   Sequence   1910 AA;


  Query Match             13.1%;  Score 170;  DB 19;  Length 1910;
  Best Local Similarity   23.8%;  Pred. No. 1.2e-05;
  Matches   57;  Conservative   31;  Mismatches   69;  Indels   34;  Gaps   10;

Qy      15 PREMISSVLIVNGKTENYILDFTPGSQASLICAVQNHIREEELIWYKEEGRVDLKSGNK- 73
           ||:: |||                 ||| || ||:|   | ::|| |||  | : |   |
Db     320 prkvkssv   . .     .... qsqvslgcsvtq todqelswyr-nqei-lnpqkev 360

Qy      74 ----INSSSVCVSSISENPNGISFICPIGPLQSVS---VSVVINVIEPPELSGNDFQIVE 126
             ||  :: :  : ::| | :: | : :|:  :   | |||    | ::|    : |
Db     361 rltginhenlimdlmvkedgg aypcfvrkdnlseqtyvpevletitpkiisatskeyys 419

Qy     127 EGSNVRLVCNVKANPQAQMMWYKNSSLLDEKSEHQIGQT--ISES---FQLSITKVEEPD 181
             | |:|||  |   : |  :   : |:   |:| |  |||      |:|:  :  |
Db     420 paepvslmcnvkqtplptitwtlddlpl lkqqsbrisqmitseqnvvsylnlsssqvrd 478

Qy     182 NGTYSCIAKSS 192
             | | | | :|
Db     479 qqvyretanns 489


RESULT   4
ABB63920
ID   ABB63920 standard; Protein; 359 AA.
XX
AC   ABB63920;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 18552.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
```

```
XX
PA   (PEKE ) PE CORP NY
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL08023.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 18552; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   359 AA;


  Query Match             13.1%;  Score 169.5;  DB 22;  Length 359;
  Best Local Similarity   27.1%;  Pred. No. 1.4e-06;
  Matches   56;  Conservative   42;  Mismatches   78;  Indels   31;  Gaps   10;

Qy       8 LLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-EEGRV 66
           ||: :|: : :     |::  |:  | |   |      |:|| : :|  :|: : :   |
Db      14 lllaifvqqtlaqrtptisyitqeqikdi--ggtvefdcsvq-yakeynvlflktdsdpv 70

Qy      67 DLKSGNKI------------NSSS--VCVSSISENDNGISFTCR--LGRDQSVSVSVVL 109
             | :|: :             |||:  : :  | | | | ::||:  :     ||  | |
Db      71 flstgstlvikdsrfslrydpnsstyklqikdiqetdag-tytcqvvistvhkvsaevkl 129

Qy     110 NVTFPPLLSGNDFQTV--EEGSNVKLVCNVKANPQAQMMWYK-NSSLLDLEKSRHQIQQT 166
           :|  ||--| |  |:|   ||| |-- |     |   - | - |---;  - -       |
Db     130 svrrppvisdnstqsvvasegsevqmecyasgyptptitwrrennailptdsa------t 183

Qy     167 SESFQLSITKVEKPDNGTYSCIAKSSL 193
                 | |  |:| | ||| |:| : :
Db     184 yvgntlriksvkkedrgtyycvadngv 210


RESULT   5
AAW94996
ID   AAW94996 standard; Protein; 431 AA.
XX
AC   AAW94996;
XX
DT   12-MAY-1999  (first entry)
XX
DE   SAF-2 polypeptide (derived from EST).
XX
KW   Sialoadhesin; SAF-2; cellular interaction molecule; cancer; inflammation;
KW   autoimmunity; allergy; asthma; rheumatoid arthritis; multiple sclerosis;
KW   cerebellar degeneration; Alzheimer's disease; Parkinson's disease; AIDS;
KW   amylotrophic lateral sclerosis; head injury; neurological abnormality;
KW   septic shock; stroke; osteoporosis; osteoarthritis; ischemia reperfusion;
KW   cardiovascular disease; kidney; myocardial; hypotension; hypertension;
KW   myelodysplastic; anaemia; infection.
XX
OS   Homo sapiens.
XX
PN   CA2226910-A.
XX
PD   02-OCT-1998.
XX
PF   26-MAR-1998;   98CA-2226910.
XX
PR   02-APR-1997;   97US-0041886.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Erickson-Miller CL,  Kikly KK;
XX
DR   WPI; 1999-106564/10.
XX
PT   New sialoadhesin (SAF-2) polypeptide and polynucleotide - useful as
PT   diagnostic reagents and for prevention and treatment of cancer,
PT   AIDS, and Alzheimer's disease
XX
PS   Claim 14; Page 23-24; 38pp; English.
XX
CC   The invention relates to a sialoadhesin polypeptide (SAF-2) which is
CC   utilised as a cellular interaction molecule. SAF-2 polynucleotides and
CC   polypeptides are useful for diagnosing susceptibility to diseases and for
CC   screening for antagonists, agonists. They can also be used to treat
CC   diseases associated with a SAF-2 imbalance, including inflammation,
CC   autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,
CC   cerebellar degeneration, Alzheimer's disease, Parkinson's disease,
CC   multiple sclerosis, amylotrophic lateral sclerosis, head injury damage
CC   and other neurological abnormalities, septic shock, cancer, stroke,
CC   osteoporosis, osteoarthritis, ischemia reperfusion injury, kidney, liver,
CC   cardiovascular diseases, myocardial infarction, AIDS, hypotension,
CC   hypertension, myelodysplastic syndromes and other abnormalities, aplastic
CC   anaemia, male pattern baldness, and bacterial, fungal, protozoan and
CC   viral infections. SAF-2 polypeptides can be used as vaccines and in gene
CC   therapy to prevent the diseases. SAF-2 polypeptides are useful for
CC   mapping genes to chromosomes, allowing gene inheritance to be studied
CC   through linkage analysis.
XX
SQ   Sequence   431 AA;


  Query Match             13.1%;  Score 169.5;  DB 20;  Length 431;
  Best Local Similarity   25.6%;  Pred. No. 1.8e-06;
  Matches   63;  Conservative   53;  Mismatches   92;  Indels   59;  Gaps   11;

Qy      16 REMISSVLFVNGKTENYILDT TPGSQASLICAVQNHTREEELLWYREEGRVDLKS -GN 72
           ::::  |  :  : :  || |   |   :| |:|          |  ::|   : |  |
Db     144 kqlsvfvtalthrpdililqtlesghsrnltcsvp---------wackqgtppmiswiga 194

Qy      73 KINS-------SSVCVSSISENDNGISFTCRL---GRDQSVSVSVVLNVTFPP------- 115
            ::|       |||   :    |:| | ||::   |   : : :| |:|::||
Db     195 svsspgpttarssvltltpkpqdhgtsltcqvtlpgtgvtttstvrldvsyppwnltmtv 254

Qy     116 ---------LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQT 166
                    |      :| || :::||| | :|| |:: | : |  | |  ||     :
Db     255 tqgdatastalgngssisvleqgslrlvcavnsnppariswtrgs--ltlcpsr-----s 307

Qy     167 SESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPIIAACV--- 221
           |    | : :|   | | ::| |:::  :: :   | ::::  |   |:  :  | |
Db     308 snpgllelprvhvrdegeftcraqnaqgsqhislslslqnegtqtsrpvsqvtlaavqga 367

Qy     222 ----VIFLTLCFGLIARRKKIMKLCMK 244
                : ||: |   |     |:: | |
Db     368 gatalatlsfcllfi-----ivrscrk 389


RESULT   6
AAW94995
ID   AAW94995 standard; Protein; 431 AA.
XX
AC   AAW94995;
XX
DT   12-MAY-1999  (first entry)
XX
DE   SAF-2 polypeptide.
XX
```

```
KW   Sialoadhesin; SAF-2; cellular interaction molecule; cancer; inflammation;
KW   autoimmunity; allergy; asthma; rheumatoid arthritis; multiple sclerosis,
KW   cerebellar degeneration; Alzheimer's disease, Parkinson's disease; AIDS;
KW   amylotrophic lateral sclerosis; head injury; neurological abnormality;
KW   septic shock; stroke; osteoporosis; osteoarthritis; ischemia reperfusion;
KW   cardiovascular disease; kidney; myocardial; hypotension, hypertension,
KW   myelodysplastic; anaemia; infection.
XX
OS   Homo sapiens.
XX
PN   CA2226910-A.
XX
PD   02-OCT-1998.
XX
PF   26-MAR-1998;   98CA-2226910.
XX
PR   02-APR-1997;   97US-0041886
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Erickson Miller CL,  Kelly KK,
XX
DR   WPI; 1999-106564/10
DR   N-PSDB; AAX17772.
XX
PT   New sialoadhesin (SAF-2) polypeptide and polynucleotide - useful as
PT   diagnostic reagents and for prevention and treatment of cancer,
PT   AIDS, and Alzheimer's disease
XX
PS   Claim 1; Page 23; 38pp; English.
XX
CC   This represents a sialoadhesin polypeptide (SAF-2) which is utilised as
CC   a cellular interaction molecule. SAF-2 polynucleotides and polypeptides
CC   are useful for diagnosing susceptibility to diseases and for screening
CC   for antagonists, agonists. They can also be used to treat diseases
CC   associated with a SAF-2 imbalance, including cancer, inflammation,
CC   autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,
CC   cerebellar degeneration, Alzheimer's disease, Parkinson's disease,
CC   multiple sclerosis, amylotrophic lateral sclerosis, head injury damage
CC   and other neurological abnormalities, septic shock, sepsis, stroke,
CC   osteoporosis, osteoarthritis, ischemia reperfusion injury, kidney, liver,
CC   cardiovascular diseases, myocardial infarction, AIDS, hypotension,
CC   hypertension, myelodysplastic syndromes and other abnormalities, aplastic
CC   anaemia, male pattern baldness, and bacterial, fungal, protozoan and
CC   viral infections. SAF-2 polypeptides can be used as vaccines and in gene
CC   therapy to prevent the diseases. SAF-2 polypeptides are useful for
CC   mapping genes to chromosomes, allowing gene inheritance to be studied
CC   through linkage analysis.
XX
SQ   Sequence   431 AA;

  Query Match          13.1%;  Score 169.5;  DB 20;  Length 431;
  Best Local Similarity;   28.6%;  Pred. No. 1.8e-06;
  Matches   63;  Conservative   54;  Mismatches   92;  Indels   59;  Gaps   11;

Qy      16 REMTSSVLTVNGKTENYILDT-TPGSQASLICAVQNHTREEELLWYREEGRVDLKS--GN 72
           ::::  |  :  : :  || |   |   :| |:|          |  ::|   :  |  |
Db     144 kqlsvivtaithrpdililqtlesqhsrallcsvp      ----wackqgtppmiswiga 194

Qy      73 KINS-------SSVCVSSISENDNGISFICRL---GRDQSVSVSVVLNVTFPP------- 115
             ::|       |||   :    |:| | ||::   |   : : :| |:|:::||
Db     195 svsspqpttarssvltltpkpqdhgtsltcqvtipgtgvtttstvrldvsyppwnltmtv 254

Qy     116 ---------LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQT 166
                    |      :| || :::||| | :|| |:: | : |  | |  ||     :
Db     255 fqpdatastalqngssslvlggslrlvcsavnsnppartswtrgs--ltlcpsr-----s 307

Qy     167 SESFQLSITKVEKEDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPIIAACV--- 221
           |    | : :|   | | ::| |:::  :: :   | :::  |   |:  :  | |
Db     308 snpgllelprvhvrdegeftcraqnaqgsqhislslslqnegtgtsrpvsqvtlaavgga 367

Qy     222     VIFLILCFGLIAKRKKIMKLCMK 244
                :  :|: |   |     |:: | |
Db     368 qatalaflsfciifi     ivrserk 389


RESULT   7
AAY88565
ID   AAY88565 standard; protein; 848 AA.
XX
AC   AAY88565;
XX
DT   07-AUG-2000  (first entry)
XX
DE   Human NCAM 140kD isoform precursor amino acid sequence.
XX
KW   NCAM; neural cell adhesion molecule; Ig1; immunoglobulin domain 1;
KW   neurite outgrowth promoter; proliferation; nerve damage; sclerosis;
KW   impaired myelination; stroke; Parkinson's disease; memory; schizophrenia;
KW   Alzheimer's disease, diabetes mellitus; circadian clock, nephrosis;
KW   treatment; prosthetic nerve guide; treatment, nervous system.
XX
OS   Homo sapiens.
XX
PN   WO200018801-A2.
XX
PD   06-APR-2000.
XX
PF   24-SEP-1999;   99WO-DK000500.
XX
PR   29-SEP-1998;   98DK-0001232.
PR   29-APR-1999;   99DK-0000592.
XX
PA   (RONN/) RONN L C B.
PA   (BOCK/) BOCK E.
PA   (HOLM/) HOLM A.
PA   (OLSE/) OLSEN M.
PA   (OSTE/) OSTERGAARD S.
PA   (JENS/) JENSEN P H.
PA   (POUL/) POULSEN F M.
PA   (SOPO/) SOPOKA V.
PA   (KALE/) KALETS I.
PA   (BERE/) BEREZIN V.
XX
PI   Ronn LCB,  Bock E,  Holm A,  Olsen M,  Ostergaard S,  Jenson PH;
PI   Poulsen FM,  Sopoka V,  Kalets I,  Berezin V;
XX
DR   WPI; 2000-293111/25.
XX
PT   Compositions that bind neural cell adhesion molecules useful for
PT   treating disorders of the nervous system and muscles e.g. Alzheimer's
PT   and Parkinson's diseases -
XX
PS   Disclosure; Fig 17; 119pp; English.
XX
CC   This sequence represents the human neural cell adhesion molecule (NCAM)
CC   amino acid sequence. NCAM is found in three forms, two of which are
CC   transmembrane forms, while the third is attached via a lipid anchor to
CC   the cell membrane. All three NCAM forms have an extracellular structure
CC   consisting five immunoglobulin domains (Ig domains). The Ig domains are
CC   numbered 1 to 5 from the N terminal. The invention relates to a compound
CC   containing a peptide which binds to the NCAM Ig1 domain. The compound
CC   binds to NCAM-Ig1/Ig2 domains, and is capable of stimulating or promoting
CC   neurite outgrowth from NCAM presenting cells, and is also capable of
CC   promoting the proliferation of NCAM presenting cells. The compound may be
CC   used in the treatment of normal, degenerated or damaged NCAM presenting
CC   cells. The compound may in particular be used to treat diseases of the
CC   central and peripheral nervous systems such as post operative nerve
CC   damage, traumatic nerve damage, impaired myelination of nerve fibres,
CC   conditions resulting from a stroke, Parkinson's disease, Alzheimer's
CC   disease, dementias, sclerosis, nerve degeneration associated with
CC   diabetes mellitus, disorders affecting the circadian clock or
CC   neuro-muscular transmission and schizophrenia. Conditions affecting the
CC   muscles may also be treated with the compound, such as conditions
CC   associated with impaired function of neuromuscular connections
```

```
CC   (e.g. genetic or traumatic shock or traumatic atrophic muscle disorders).
CC   Conditions of the gonads, pancreas (e.g. diabetes mellitus types I and
CC   II), kidney (e.g. nephrosis), heart, liver and bowel may also be treated
CC   using the compound. The compound is used in a prosthetic nerve guide, and
CC   also to stimulate the ability to learn, and to stimulate the memory of a
CC   subject.
XX
SQ   Sequence   848 AA;


  Query Match             12.8%;  Score 166;  DB 21;  Length 848;
  Best Local Similarity   28.0%;  Pred. No. 9e-06;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      41 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 86
           | :| |              : :| : ||: |    :| : ::|  ::  ||: . ||
Db     324 qvtltceasgdpipsitwrtstrnisseektl----dghmvvrsharv--ssltlksiqy 377

Qy      87 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 142
            | |  : |     :|:|   | |: | | : | | |       ||: | : | | | |
Db     378 tdag-eyictasntigqd---sqsmylevqyapklqgpvavytwegnqvnitcevfayps 433

Qy     143 AQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 201
           | : |::: ||     | :| |  :  | :|   : | | |:| | . .  ;|.|
Db     434 atiswfrdgqllpssnysnikiyntpsasylevtpdsendfgnynctavnrigqeslefi 493

Qy     202 LIVKDKTVGVP----IEPIIAACVVIF 224
           |:  | |   |    :|| :   | |
Db     494 lvqad-tpsspsidqvepysstaqvqf 519


RESULT   8
AAB61142
ID   AAB61142 standard; Protein; 404 AA.
XX
AC   AAB61142;
XX
DT   30-MAR-2001  (first entry)
XX
DE   Human NOV12 protein.
XX
KW   Human; NOVX; antiinflammatory; cytostatic; neuroprotective;
KW   cerebroprotective; immunomodulator; vulnerary; vasotropic; gene therapy;
KW   hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
KW   diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
XX
OS   Homo sapiens.
XX
PN   WO200075321-A2.
XX
PD   14-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15303.
XX
PR   03-JUN-1999;   99US-0137322.
PR   16-MAR-2000; 2000US-0189810.
PR   22-MAR-2000; 2000US-0191158.
PR   30-MAR-2000; 2000US-0193086.
PR   31-MAY-2000; 2000US-0137322.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Fernandes E,  Herrman J,  Vernet C;
XX
DR   WPI; 2001-102403/11.
DR   N-PSDB; AAF27860.
XX
PT   New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
PT   diagnostic marker, protein therapeutic, antibody or small molecule drug
PT   target for treating immune, proliferative and metabolic diseases and
PT   wound healing
XX
PS   Claim 1; Page 44-46; 194pp; English.

XX
CC   The present sequence is a new isolated polypeptide (NOVX). The NOVX
CC   polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
CC   treating or preventing NOVX-associated disorders. They are also useful
CC   for determining the presence of or a predisposition to a disease
CC   associated with altered levels of the NOVX polypeptide or nucleic acid.
CC   These NOVX-associated disorders include hyperplasias, tumours,
CC   restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
CC   rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
CC   oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
CC   are especially useful in gene therapy. Specifically, NOVX is useful as
CC   a diagnostic marker or prognostic marker, protein therapeutic and
CC   antibody target or small molecule drug target to treat disorders in the
CC   immune response pathway, thyroid and metabolic diseases, bone metabolic
CC   disorders, diseases of the pancreas (e.g. diabetes or digestive
CC   disorders), proliferative diseases, or tissue regeneration and
CC   development (e.g. wound healing or treatment of burns).
XX
SQ   Sequence   404 AA;


  Query Match             12.5%;  Score 162;  DB 22;  Length 404;
  Best Local Similarity   22.4%;  Pred. No. 7.6e-06;
  Matches   64;  Conservative   48;  Mismatches  112;  Indels   62;  Gaps   13;

Qy      14 LPREMTSSVLTVNG-----KTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRV-D 67
           :| : : : ||| |      :   :       |    | |          :: |:: :  : |
Db     118 mpvktskayltvlgvpgkpqisgfsspvmegdlmqltcktsgskpaadirwfkndkeikd 177

Qy      68 LK--SGNKINSSSVCVSS-----ISENDNGISFTCRLGRDQ----------------SVS 104
           :|       |  :  |||      :  :|:|::  ||:  :                  : |
Db     178 vkylkeedanrktftvsstldfrvdrsddgvavicrvdheslnatpqvamqvleihytps 237

Qy     105 VSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQ-MMWYKN-SSLLDLEKSRHQ 162
           | :: :  ||           :||  : | |  |  |  : ::| |:   | | ::
Db     238 vkiipstpfp           qeqqpliltceskqkplpepvlwtkdgqelpdpdr---- 282

Qy     163 IQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTV----GVPIEPIIA 218
                 :|:|  : | |||||  | | :::   | :: ||| |          |    :|
Db     283 - mvvsgrelnilflnktdngtyrceatntigqssaeyvlivhdpnalaqqngpdhaliq 340

Qy     219   ACVVIFLTLC   -FGLIARRKKIMKLCMKDKD    PHSETAL 254
               ||:|:|||     |    | |   |  : |     | ::||:
Db     341 givavvvfvtlcsifllgrylarhkgtyltneakgaedapdadtai 386


RESULT   9
ABB71337
ID   ABB71337 standard; Protein; 1079 AA.
XX
AC   ABB71337;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 40803.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
```

```
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15440.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 40803; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840 ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence   1079 AA;


  Query Match             12.4%;  Score 161;  DB 22;  Length 1079;
  Best Local Similarity   24.3%;  Pred. No. 3.4e-05;
  Matches   46;  Conservative   34;  Mismatches   81;  Indels   28;  Gaps    6;

Qy      23 LTVNGKTENYILDFTPGSQASLICAVQNHTREEELLWYR-----EEGRVDL--------- 68
           : :||   :    |  ||:  | | :        | ||        |  ::
Db     265 IdlnqvkhH----tvqqskvvltcdihqarpavnltwyntttiissgeneltevrskslc 320

Qy      69 KSGNKINSSSVCVSSISPENNGISFTC-------RTGPDQSVSVSVVLNVTFPPLLSGND 121
           ||    :: |  : : :  :|   | |       :: |:: :| :: | | :||:  :
Db     321 kvdqrthtqscliftvtrkddrstyscnivlginrekpissaltlsvlyppvvkvsp 380

Qy     122 FQIVEEGSNVKLV CNVKANPQA--QMMWYKNSSLLDLEKSRHQIQQTSESPQLSIIKVR 178
           | : |: |   ||| :  |: ||:|  |:::  : |     ||:  | |   |
Db     381 saitantselvllecfyaapasipqvewfrndilnnndthqkqgnsenvalvikatc 440

Qy     179 KPDNGTYSC 187
           | | | |||
Db     441 kedlqnysc 449


RESULT  10
AAY33741
ID   AAY33741 standard; Protein; 444 AA.
XX
AC   AAY33741;
XX
DT   09-NOV-1999  (first entry)
XX
DE   Beta-secretase.
XX
KW   Beta-secretase; beta amyloid protein precursor, APP; Down's syndrome;
KW   Alzheimer's disease
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Signal_peptide
FT   Protein         24..444
FT                   /label= beta-secretase
FT   Region          377..399
FT                   /note= "Putative transmembrane region"
XX
PN   US5942400-A.
XX
PD   24-AUG-1999.
XX
PF   07-JUN-1996;   96US-0659984.
XX
PR   07-JUN-1996;   96US-0659984.
PR   07-JUN-1995;   95US-0480498.
PR   07-JUN-1995;   95US-0485152.
XX
PA   (ELAN-) ELAN PHARM INC.
XX
PI   Anderson JP, Jacobson-Croak KL, Sinha S;
XX
DR   WPI; 1999-517417/43.
DR   N-PSDB; AAZ06640.
XX
PT   A method for detecting human beta secretase cleavage of polypeptides
PT   useful for identifying beta-secretase inhibitors
XX
PS   Examples; Fig 1; 43pp; English.
XX
CC   This sequence is the human beta secretase enzyme polynucleotide.
CC   Beta secretase is capable of cleaving the beta-amyloid precursor protein
CC   (APP) (AAY33742; swedish mutant APP). This enzyme is used in a method for
CC   detecting human beta-secretase cleavage of polypeptides and for
CC   identifying beta-secretase inhibitors. The protein has a deduced
CC   molecular weight of 49 kD, although this is without glycosylation and
CC   the expected molecular weight is therefore higher. Inhibition of
CC   beta secretase activity would be useful for chemical modelling of a
CC   critical event in the pathology of Alzheimer's disease. Inhibitors of
CC   beta secretase would be useful for the prevention and treatment of
CC   Alzheimer's disease and Down's Syndrome.
XX
SQ   Sequence   444 AA;


  Query Match             12.2%;  Score 158;  DB 20;  Length 444;
  Best Local Similarity   20.6%;  Pred. No. 1.9e-05;
  Matches   67;  Conservative   48;  Mismatches  109;  Indels  132;  Gaps   13;

Qy      14 LPKEMISSVLIVNGKIE-----NYILDFTPGSQASLICAVQNHTREEELLWYREEGRV-D 67
           :| : : : |||  |  |     :     |   | |           :: ||: :  : |
Db     178 mpvltsha,l*clqqpekpqisisspenegdlmqlldktcpskpsadirwfkndkeikd 177

Qy      68 LK--SGNKINSSSVCVSS-----ISENDNGISFTCKIGKDQ----------------SVS 104
           :|      |  :  |||      :  :|:|::  ||:  :                  : |
Db     178 qkylkeedanrk*ftvsstldfrvdrsddqqvaicrvdhesiatpqpqamqyleihvtps 237

Qy     105 VSVVLNVTFPPLLSGNDFQIVEEGSNVKLVCNVKANPQAQ-MMWYKN-SSLLDLEKSRHQ 162
           | :: :  ||           :||  : | |  |  |  : ::| |:   | | ::
Db     238 vkilpstpfp           qeqqpliltceskgkplpepvlwtkdqqelpdpdr---- 282

Qy     163 IQQTSESPQLSIIKVRKPDNGTYSCIAKSSLGKIESIDFHLIVKD-------- -- --- 206
             :|:|  : | |||||  | | :::   | :: ||| |
Db     283   mvvsqrelnilflnktdnqtyrccatntiqqssaeyvlivkdvpntllptfipslttn 340

Qy     207 ----KTVGVPIEPIIAA-----------------------CVVIPLTIC----PGLI 232
                 || :   |  :|                       |||:|||     |
Db     341 atvtttvaittspttsattssirdpnalaqqnqpdhaliqqivavvvivtlcsiflqry 400

Qy     233 APPKFIMKLQMKDKD----PHSETAL 254
           | |   |  : |     | ::||:
Db     401 larhkgtyltneakqaedapdadtai 426


RESULT  11
AAB47251
ID   AAB47251 standard; Protein; 444 AA.
XX
AC   AAB47251;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Beta-secretase.
XX
KW   Beta secretase, isotype, beta amyloid precursor protein; APP;
```

```
KW   beta-amyloid peptide; beta-AP; Alzheimer's disease; Downs syndrome;
KW   HCHWA-D; Swedish mutation.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /note= "Putative signal peptide"
FT   Protein         24..444
FT                   /note= "Mature beta-secretase"
FT   Peptide         125..141
FT                   /label= Peptide 1
FT   Peptide         190..200
FT                   /label= Peptide 5
FT   Peptide         240..259
FT                   /label= Peptide 2
FT   Peptide         261..272
FT                   /label= Peptide 3
FT   Domain          377..399
FT                   /note= "Putative transmembrane domain"
XX
PN   US6221645-B1.
XX
PD   24-APR-2001.
XX
PF   07-JUN-1996;   96US-0660531.
XX
PR   07-JUN-1995;   95US-0480498.
XX
PA   (ELAN-) ELAN PHARM INC.
XX
PI   Chrysler SMS,  Sinha S,  Keim PS,  Anderson JP,  Tan H,  McConlogue LC;
XX
DR   WPI; 2001-315578/33.
DR   N-PSDB; AAC85809.
XX
PT   Novel antibody that specifically binds native beta-secretase protein,
PT   useful for raising anti-idiotypic antibodies and for detecting or
PT   diagnosing pathological conditions related to presence of respective
PT   antigens -
XX
PS   Example; Fig 1; 42pp; English.
XX
CC   This sequence represents beta-secretase protein which is characterized
CC   by an ability to cleave the 695-amino acid isotype of beta-amyloid
CC   precursor protein (APP) between amino acids 596 and 597.  Beta-
CC   secretase is thought to be responsible for the pathogenic processing
CC   of APP to form beta-amyloid peptide (beta-AP) in beta-AP related
CC   conditions, e.g. Alzheimer's disease, Downs syndrome, HCHWA-D etc.
CC   Beta-secretase has a molecular weight of 260-300 kD and will bind
CC   to wheat germ agglutinin but not to concanavalin A.  Beta-secretase
CC   will cleave both the wild type and the Swedish mutation of APP.
XX
SQ   Sequence   444 AA;
```

```
  Query Match           12.2%;  Score 158;  DB 22;  Length 444;
  Best Local Similarity   20.6%;  Pred. No. 1.9e-05;
  Matches   67;  Conservative   48;  Mismatches  109;  Indels  102;  Gaps   13;

Qy      14 LPREMTSSVLTVNGKTE-----NYILDTTPGSQASLICAVQNHTREEELLWYPEEGRV-D 67
           :| : : : ||| |  |       :     |    | |          :: |:: :  : |
Db     118 mpvktskayltvlgvpekpqisgfsspvmegdlmqltcktsgskpaadirwfkndkeikd 177

Qy      68 LK--SGNKINSSSVCVSS-----ISENDNGISFTCRLGRDQ----------------SVS 104
           :|       |  :  |||     :  :|:|::  ||:  :                  : |
Db     178 vkylkeedanrktftvsstldfrvdrsddgvavicrvdheslnatpqvamqvleihytps 237

Qy     105 VSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQ-MMWYKN-SSLLDLEKSRHQ 162
           | :: :  ||           :||  : | |  |  |  : ::| |:   | | ::
Db     238 vkiipstpfp----------qeqqpiiitceskgkplpepvlwtkdggelpdpdr---- 282

Qy     163 IQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKD---------------- 206
```

```
                :|:|  : | |||| | | :::   | :: ||| |
Db     283 --mvvsgrelnilflnktdngtyrceatntiqqssaeyvlivhdvpntllptiipsltt 340

Qy     207 ----KTVGVPIEPIIAA-----------------------CVVIFLTLC----FGLI 232
               || :   |  :|                        ||:|:|||     |
Db     341 atvtttvaittsprtsartssirdpnalaqqngpdhaiiggivavvvtvticsiflgry 400

Qy     233 ARRKKIMKLCMKDKD----PHSETAL 254
           | |   |  : |     | ::||:
Db     401 larhkgtyltneakgaedapdadtai 426
```

```
RESULT  12
AAB50692
ID   AAB50692 standard; Protein; 1415 AA.
XX
AC   AAB50692;
XX
DT   19-MAR-2001  (first entry)
XX
DE   C. elegans UNC-40 protein SEQ ID NO:93.
XX
KW   Human; Caenorhabditis elegans; UNC-5; splice variant; nematode worm;
KW   protein protein interaction; identification.
XX
OS   Caenorhabditis elegans.
XX
PN   WO200073328-A2.
XX
PD   07-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-EP05108.
XX
PR   01-JUN-1999;   99GB-0012755.
XX
PA   (DEVG-) DEVGEN NV.
XX
PI   Van Criekinge W,  Roelens I,  Bogaert T,  Verwaerde P;
XX
DR   WPI; 2001-016508/02.
DR   N-PSDB; AAC90960.
XX
PT   Three variants of human unc-5C cDNAs (unc-5Cb, unc-5Cc and unc-5C8) and
PT   a human unc-5HS1 cDNA, useful in yeast two hybrid experiments for
PT   identifying unknown human cDNAs which encode proteins that interact
PT   with the human unc-5C protein -
XX
PS   Disclosure; Page 233-238; 246pp; English.
XX
CC   The present invention describes 3 variants of human unc-5C cDNAs
CC   (unc-5Cb, unc-5Cc and unc-5C8) which correspond to alternatively spliced
CC   unc-5C transcripts, and a human unc-5HS1 cDNA which shares homology with
CC   the Rattus norvegicus unc-5HS1 cDNA. Also described are assays based on
CC   protein-protein-interactions between the unc-5 protein and a variety of
CC   different interacting proteins. The unc-5C variant cDNAs and unc-5HS1
CC   cDNA are useful in methods for identifying compounds which reduce or
CC   inhibit the lethal phenotype associated with the expression of the
CC   unc-5 death domain in yeast. They are also useful in yeast two hybrid
CC   experiments for identifying unknown human cDNAs which encode proteins
CC   that interact with the human unc-5C protein. AAC90914 to AAC90971 and
CC   AAB50646 to AAB50693 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   1415 AA;
```

```
  Query Match           12.2%;  Score 157.5;  DB 22;  Length 1415;
  Best Local Similarity   27.8%;  Pred. No. 9.8e-05;
  Matches   58;  Conservative   35;  Mismatches   83;  Indels   33;  Gaps   10;

Qy      16 REMTSSVLTVNGKT--ENYILDTTP-------GSQASLICAVQNHTREEELLWYREEGR- 65
           |   :: |||  :|     :  |||       | :  | | | :  |  :: | ::  :
Db     227 rssqtarltvttetvsnelvfittprlqvveqgdeflleclvaslir-pqvrwlkdsrqi 285
```

```
Qy      66 -VDLKSGNKINSSSVECVGSIENGNGTSFDCPLGSRPQSVSVSVVLNVTEPPLLSGNKFQ 125
           ||      ::  ||: ||  |  | |: :|||    | |:  :| : |  || ::
Db     286 lvdqvrirrqrssilzsrssicdrql ytcrasnddsidravsevevrappritirptt 344

Qy     126 TVE-EGSNVELVCNVFANPQAQMMWYKN------SSLLDLEKSRHQIQQTSESFQLSITKVEKEDNGTYSCIAKSS 192
           |  | ::|:| |    | |:||: ||||      |   :| :|           | |
Db     345 kqavcsdgvtwrpdlylrnvtpqdagkyhcqafdtgsgyetvtlepqr  --        lril 393

Qy     193 L 193
           |
Db     394 q 394

RESULT  13
ABB63044
ID   ABB63044 standard; Protein; 467 AA.
XX
AC   ABB63044;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 15924.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   24-MAR-2000; 2000US-191647P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWO,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07147.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 15924; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737 ABB72972).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   467 AA;


  Query Match             12.1%;  Score 157;  DB 22;  Length 467;
  Best Local Similarity   26.5%;  Pred. No. 2.5e-05;
  Matches   48;  Conservative   37;  Mismatches   72;  Indels   24;  Gaps    8;

Qy      24 IVNGKTLNYTLDITEGSQASLQCAVQNHTKEEGLLWYKEEGKVVDLKSG--NKINSSSVCV 81
           |:     |  :    ||  :| |    :     : |:::    |: ||  :  :|||: :
Db     146 tlralphnqqvtarkqstvtleckasqnp-vptitwfkk----dvfsgpthlsdsstlil 190

Qy      82 SSTSENLNCTSEIDPLI--SPQQSVSVSVVLNVTEPPLISGNKFQIVEG--      FGSNVK 132
           ::  :  | :: |    |    ||: : | :  || :       |||        || :|:
Db     191 engvdrhhaq-tyqesadnqykdrvsmdiqlttlsppei------tvekswvhaseqydvc 243

Qy     133 LVCNVFANPQAQMMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKEDNGTYSCIAKSS 192
           ||| |  :   ::|:|||:|| |||    |  :    : : | |    :  | | |||:| ::
Db     244 lvcivhqdvnsemlwygnstlld atdrrsmyprddrysliirnfqptdtgnyscvadna 302

Qy     193 L 193
           |
Db     303 l 303

RESULT  14
AAB70070
ID   AAB70070 standard; Protein; 595 AA.
XX
AC   AAB70070;
XX
DT   14-MAY-2001  (first entry)
XX
DE   Human secreted protein #9.
XX
KW   Human; secreted protein; immunomodulatory; antisclerotic;
KW   dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW   immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW   ophthalmological, neuroprotectant, nootropic, anticonvulsant, vaccine;
KW   antialzheimers; antiparkinsonian, antimicrobial, vulnerary; gene therapy;
KW   immune disorder, hyperproliferative, cardiovascular, angiogenic;
KW   neurological; infection.
XX
OS   Homo sapiens.
XX
PN   WO200112776-A2.
XX
PD   22-FEB-2001.
XX
PF   15-AUG-2000; 2000WO-US22350.
XX
PR   16-AUG-1999;   99US-0148759.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Shi Y,  Young PE,  Ebner R,  Soppet DR,  Ruben SM;
XX
DR   WPI; 2001-244245/25.
DR   N-PSDB; AAF76851.
XX
CT   Nucleic acids encoding 18 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Claim 11; Page 364-366; 380pp; English.
XX
CC   The present sequence is one of 18 novel human secreted proteins. The
CC   nucleic acids encoding the proteins and the proteins themselves may be
CC   used in the prevention, diagnosis and treatment of diseases including
CC   immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
CC   and human immunodeficiency virus (HIV) infections), hyperproliferative
CC   disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC   (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC   arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC   neovascularisation and diabetic retinopathy), neurological disorders
CC   (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC   infectious diseases and/or for promoting wound healing, regeneration
CC   and/or chemotaxis. The nucleic acid molecules may also be used as DNA
CC   probes in diagnostic assays to detect and quantitate the presence of
CC   similar nucleic acid sequences in samples. The polypeptides may also be
CC   used as antigens in the production of antibodies and in assays to
CC   identify modulators of protein expression and activity.
XX
SQ   Sequence   595 AA;
```

```
  Query Match             12.0%;  Score 155;  DB 22;  Length 595;
  Best Local Similarity   23.2%;  Pred. No. 5.1e-05;
  Matches   63;  Conservative   46;  Mismatches   84;  Indels   78,  Gaps   12.

Qy      31 NYILD---------------TTPGSQAS-----LICAVQNHTREEELLWYKEEGK----- 65
           ||| |                 : ||:  |     | |:|          |  |:|
Db     258 nyiydklsvhvtalthmptfsipqtlesghprnltcsvp---------waceqgtpptit 308

Qy      66 ---VDLKSGNKINSSSVCVSSISE-NDNGISFTCRL---GRDQSVSVSVVLNVTFPP--- 115
              : | :   : |  :| | :  |:| | ||::   |   ::: :| ||:::||
Db     309 wmgasvssldptitrssmlslipqpqdhgtsltcqvtlpgagvtmtravrlnisyppqnl 368

Qy     116 ------------LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQ 162
                        |      :| || :: ||| | :|| |:: |   |  | | |
Db     369 tmtvfqqdgtasttlrngsalsvlegqslhlvcavdsnppar1swtwgs--ltlsps--- 423

Qy     163 IQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPI---- 216
             |:|    | : :|   | | ::| |:: | :: :   | ::::  |   ||  :
Db     424 --qssnlgvlelprvhvkdegeftcraqnplgsqhislslslqneytgkmrpisgvtlga 481

Qy     217 ---IAACVVIFLTLCFGLIARRKKIMKLCMK 244
               |  ::||  |   :     ::: | |
Db     482 fggagatalvflyfciifv-----vvrscrk 507


RESULT  15
AAG77857
ID   AAG77857 standard; Protein; 537 AA.
XX
AC   AAG77857;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Mutant protein mMuSK-RI.
XX
KW   Mutated muscle specific tyrosine kinase receptor; selectable marker;
KW   hMuSK-R; human; mMuSK-R; antibody; mutant; mutein; immunoselection.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "Signal peptide of hMuSK-R"
FT   Domain          20..493
FT                   /note= "Extracellular domain"
FT   Domain          494..515
FT                   /note= "Transmembrane domain"
XX
PN   WO200172834-A1.
XX
PD   04-OCT-2001.
XX
PF   28-MAR-2001; 2001WO-EP03543.
XX
PR   30-MAR-2000; 2000US-0539248.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Pippig SD,  Veres G;
XX
DR   WPI; 2001-626257/72.
XX
PT   Identifying genetically modified mammalian cells comprising a nucleic
PT   acid sequence encoding a mutated muscle-specific tyrosine kinase
PT   receptor provides a selectable marker for transformation  -
XX
PS   Claim 5; Page-; 52pp; English.
XX
CC   The sequence represents the mutant protein mMuSK-RI. The invention
CC   concerns a method of using a muscle specific tyrosine kinase receptor
CC   molecule (MuSK-R) or a mutated MuSK-R (mMuSK-R) thereof as a selectable
CC   marker in mammalian cells, particularly human cells. Preferred markers
CC   are mMuSK-R's incapable of signal transduction and preferably molecules
CC   wherein the intracellular domain has been modified by deletion of the
CC   signalling region. The invention also relates to a method of identifying
CC   genetically modified mammalian cells including introducing a mMuSK-R into
CC   a target cell as a selectable marker. Further, a method for the
CC   immunoselection of transduced mammalian cells is disclosed comprising
CC   identifying the transduced cells by incubation of the cells with an
CC   antibody which recognises and binds specifically to a MuSK-R or mMuSK-R
CC   thereof.
CC   Note. The present sequence is not shown in the specification but is
CC   derived from the human wild type muscle specific tyrosine kinase receptor
CC   protein shown in Figure 2 (AAG77856).
XX
SQ   Sequence   537 AA;


  Query Match             11.9%;  Score 154.5;  DB 22;  Length 537;
  Best Local Similarity   25.2%;  Pred. No. 4.9e-05;
  Matches   56;  Conservative   41;  Mismatches   82;  Indels   43;  Gaps   11;

Qy       7 LLLVILF-----LPRE--MTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLW 59
           :| :: |     ||:   :|: : ||:   |           |: :|||:::  : |: |
Db      12 iltlvafsgteklpkapvittpletvdalveev   atfmcavesyp-qpeisw 61

Qy      60 YREEGRV---DLKSGNKINSSSVCVSSISENDNGISFTCRLGRD--------QSVSVSVV 108
           | : :   | :   : |   : : |: ::|:|| : |              :: | :
Db      62 trnkilikldtrysirengqlltilsvedsddgi ycctanngvggaveseqalqvkmk 120

Qy     109 LNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSE 168
             :| ||:       : ||    | |    ||:  : | |  | |  | ||  :    |
Db     121 pkitrppi-----nvkileglkavlpcttmgnpkpsvswikgdspl-rensriav---le 171

Qy     169 SFQLSITKVEKPDNGTYSCIAKSSL-----KTESLDFHLIVK 205
           |  | |  |:| | | | |:||:||     |   |:| :  :
Db     172 sgslrihnvqkedagqyrcvaknslgtayskvvklefevfar 213


Search completed: September 17, 2002, 13:09:55
Job time: 11322 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:09:55 ; Search time 86.46 Seconds
(without alignments)
336.587 Million cell updates/sec

Title: US-09-863-823-6
Perfect score: 1336
Sequence: 1 MAWKSSV1MQMGKFLLLV1L.....RKKIMKLCMKDKDPHSETAL 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	311	23.3	66	22	AAU18000	Human immunoglobul
2	173	12.9	359	22	ABB63920	Drosophila melanog
3	171	12.8	1571	19	AAW42087	Human Down syndrom
4	170	12.7	1910	19	AAW42086	Human Down syndrom
5	169.5	12.7	431	20	AAW94995	SAF-2 polypeptide
6	169.5	12.7	431	20	AAW94995	SAF-2 polypeptide.
7	166	12.4	848	21	AAY88565	Human NCAM 140kD i
8	162	12.1	404	22	AAB61142	Human NgV12 protei
9	161	12.1	1079	22	ABB71337	Drosophila melanog
10	158	11.8	444	20	AAY33741	Beta-secretase. H
11	158	11.8	444	22	AAB47251	Beta-secretase. H
12	157.5	11.8	1415	22	AAB50692	C. elegans UNC-40
13	157	11.8	467	22	ABB63044	Drosophila melanog
14	155	11.6	595	22	AAB70070	Human secreted pro
15	154.5	11.6	537	22	AAG77857	Mutant protein mMu
16	154.5	11.6	576	22	AAG77858	Mutant protein mMu
17	154.5	11.6	869	22	AAG77856	Protein of muscle
18	154.5	11.6	869	22	AAB68421	Amino acid sequenc
19	154	11.5	582	17	AAR92256	Neural cell adhesi
20	154	11.5	761	17	AAR92255	Neural cell adhesi
21	154	11.5	868	18	AAW26610	Rat muscle-specifi
22	154	11.5	868	18	AAW26507	Rat Dmk receptor.
23	153	11.5	467	19	AAW59994	Sialoadhesin famil
24	153	11.5	467	21	AAY97218	Sialic acid bindin
25	153	11.5	467	21	AAB25581	CD33-like 3 protei
26	153	11.5	467	22	AAM23894	Human EST encoded
27	153	11.5	467	22	AAU14345	Human novel protei
28	153	11.5	467	22	AAU14581	Human novel protei
29	153	11.5	467	22	AAU14582	Human novel protei
30	153	11.5	467	22	AAB29188	Siglec 5 protein.
31	152	11.4	322	22	AAE04367	Human kinase (PKIN
32	152	11.4	869	18	AAW26611	Human muscle-speci
33	152	11.4	869	18	AAW26506	Human Dmk receptor
34	151.5	11.3	983	21	AAY70775	Follistatin-relate
35	149.5	11.2	851	22	AAM78324	Human protein SEQ
36	149	11.2	478	17	AAR92718	Mouse muscle-local
37	149	11.2	860	17	AAR92716	Mouse muscle-local
38	149	11.2	868	17	AAR92717	Mouse muscle-local
39	149	11.2	1447	16	AAR68553	Deleted in colorec
40	149	11.2	1447	20	AAY33498	Human DCC protein.
41	149	11.2	1447	22	AAB50693	Human UNC-40 prote
42	149	11.2	1728	12	AAR13144	Deleted in Colorec
43	149	11.2	2037	22	ABB71928	Drosophila melanog
44	148.5	11.1	245	19	AAW62577	Alternatively spli
45	148.5	11.1	308	17	AAW05157	Human LAMP residue

ALIGNMENTS

RESULT 1
AAU18000
ID AAU18000 standard; Protein; 66 AA.
XX
AC AAU18000;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immunoglobulin polypeptide SEQ ID No 145.
XX
KW Immunoglobulin; signal transduction pathway protein; cancer;
KW antisense therapy; gene therapy; neurological disorder; renal disorder;
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;
KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
OS Homo sapiens.
XX
PN WO200155315-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01326.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215145.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216882.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225427.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227099.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230448.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231908.
PR 14-SEP-2000; 2000US-0232347.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234221.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239945.
PR 13-OCT-2000; 2000US-0239947.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0250748.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457725/49.
DR N-PSDB; AAS28788.
XX
PT Isolated novel immunoglobulin polypeptide for monitoring the presence
PT and progression of diseases and for diagnosis -
XX
PS Claim 11; SEQ ID No 145; 551pp; English.
XX
CC Sequences AAU17977-AAU18087 represent immunoglobulin polypeptides of the
CC invention. The polypeptides and their associated polynucleotides can be
CC used to diagnose a pathological condition or a susceptibility to a
CC pathological condition in a subject by determining the presence or
CC absence of a mutation in a DNA sequence or determining the presence or
CC amount of expression of the protein. Alternatively the identification of
CC a binding partner to a sequence allows determination of changes in

```
CC   protein activity. The sequences can be used as research tools for
CC   receptors or other signal transduction pathway proteins that interact
CC   with the polypeptides of the invention and can be used to treat, prevent
CC   or diagnose various types of disorders such as neurological disorders,
CC   cardiovascular disorders, gastrointestinal disorders, reproductive
CC   disorders, immune system disorders, renal disorders, muscular disorders,
CC   pulmonary disorders, proliferative disorders and cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   66 AA;


  Query Match             23.3%;  Score 311;  DB 22;  Length 66;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-20;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    153 MMWYKNSSLLDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIV 212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mmwyknsslldleksrhqiqqtsesfqlsitkvekpdngtysciaksslktesldfhliv 60

Qy    213 K 213
          |
Db     61 k 61


RESULT   2
ABB63920
ID   ABB63920 standard; Protein; 359 AA.
XX
AC   ABB63920;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 18552.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL08023.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 18552; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072)
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   359 AA;


  Query Match             12.9%;  Score 173;  DB 22;  Length 359;
  Best Local Similarity   27.3%;  Pred. No. 8e-07;
  Matches   80;  Conservative   43;  Mismatches   85;  Indels   32;  Gaps   11;

Qy      3 WKSSVIMQMGRFLLLVILFLPREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTRE 62
          |: |:   :   ||| | |: : :     |::  |:  | |   |      |:|| : :|
Db      2 wrpsisncvwstlllai-fvqqtlaqrtptisyitqeqikdi--ggtvefdcsvq-yake 57

Qy     63 EFLIWYP-EFGRVDLKSGNKI------------NSSS--VCVSSISFNDNGISFTCR-- 104
            :|: : :   | | :|: :              |||:  : :  | | | | ::||:
Db     58 ynvlflktdsdpvflstgstlvikdsrfslrydpnsstyklqikdiqetdag tytcqvv 116

Qy    105 LGRDQSVSVSVVLNVTFPPLLSGNDFQTV--EEGSNVKLVCNVKANPQAQMMWYK-NSSL 161
          :     ||  | |:|  ||::| |  |:|   ||| |:: |     |   : | : |:::
Db    117 istvhkvsaevklsvrrppvisdnstgsvvasegsevqmecyasqyptptitwrrennai 176

Qy    162 LDLEKSRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSL 201
          |  : :      |     | |  |:| | ||| |:| : :
Db    177 lptdsa------tyvgntlriksvkkedrgtyycvadnqv 210


RESULT   3
AAW42087
ID   AAW42087 standard; Protein; 1571 AA.
XX
AC   AAW42087;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Human Down syndrome-cell adhesion molecule DS CAM2.
XX
KW   DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;
KW   signal transduction; trisomy 21; mental retardation;
KW   holoprosencephaly; corpus callosum agenesis;
KW   schizencephaly; diagnosis; assay; human.
XX
OS   Homo sapiens.
XX
PN   WO9817795-A1.
XX
PD   30-APR-1998.
XX
PF   23-OCT-1997;   97WO-US19547.
XX
PR   25-OCT-1996;   96US-0029322.
XX
PA   (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI   Korenberg JR;
XX
DR   WPI; 1998-271791/24.
DR   N-PSDB; AAV31988.
XX
PT   New isolated Down's Syndrome-cell adhesion molecule - used to
PT   develop products for detection, diagnosis and therapy of
PT   developmental and neurological abnormalities
XX
PS   Claim 2; Page 90-95; 109pp; English.
XX
CC   This polypeptide comprises Down syndrome-cell adhesion molecule
CC   DS-CAM2, an extracellular soluble protein belonging to a novel
CC   subclass of the Ig superfamily with highest homology to neural cell
CC   adhesion molecules.  Its amino acid sequence was deduced from cDNA
CC   clones (see AAV31982) isolated from a trisomy 21 foetal brain library.
CC   It is a splice variant of membrane-bound DS-CAM1 (see AAW42086), and
CC   lacks the entire transmembrane domain of DS-CAM1.  The invention
CC   provides human and murine DS-CAM nucleic acid sequences (see also
```

```
CC   AAV41981, AAV41985-87), expression vectors and host cells, transgenic
CC   animals, antibodies, antisense oligonucleotides, and primers
CC   derived from DS-CAM nucleic acids. DS CAM polypeptides are associated
CC   with developmental and neurological processes. They can be used in
CC   e.g. neural prosthetic devices used in entubulation methods of
CC   repairing (regenerating) damaged or severed peripheral nerves, and
CC   also in bioassays to identify agonists and antagonists. The products
CC   can also be used in detection, diagnosis and therapy of developmental
CC   and neurological abnormalities such as Down syndrome, mental
CC   retardation, holoprosencephaly, agenesis of the corpus callosum,
CC   or schizencephaly.
XX
SQ   Sequence   1571 AA;


  Query Match          12.8%;  Score 171;  DB 19;  Length 1571;
  Best Local Similarity   30.4%;  Pred. No. 8.4e-06;
  Matches   58;  Conservative   39;  Mismatches   69;  Indels   34;  Gaps   10;

Qy      24 PREMTSSVLTVNGKTENYTLDFTPGSQASLTCAVQNETREEELLWYREEGKVDLKSGNK- 81
           ||:: |||                  ||| || |:|   | ::|| |||  | : |  |
Db     320 prkvkssv----------------qsqvslscsvtq-tedqelswyr-nqei-lnpgknv 360

Qy      82   --TNSSSVCVSSISENDNGTSFICTLTGRDQSVS---VSVVCNVTFPPPTSGNDFQTVE 134
             ||  :: :  : ::| | :: | : :|:  :    | |||    | ::|    : |
Db     361 ritqinhenlimdhmvksdqq-ayqetvrkdklsqqdyvqvvledqtpkilsafsekvvs 419

Qy     135 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRRQTQQ--TSESFQLS---TTKVEKND 189
           | |:|||  |   : |  :   : |:   |:| |  |||   ||   |:  :  |
Db     420 paepvslmcnvkqtplptitwtlddppi lkqqshrisqmitseqnvlsylnlsssqvrd 478

Qy     190 NGTYSCIAKSS 200
             | | | | :|
Db     479 qqvyretanns 489


RESULT   4
AAW42086
ID   AAW42086 standard; Protein; 1910 AA.
XX
AC   AAW42086;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Human Down syndrome cell adhesion molecule DS CAM1.
XX
KW   DS-CAM1; Down syndrome cell adhesion molecule; neural cell;
KW   signal transduction; trisomy 21; mental retardation;
KW   holoprosencephaly, corpus callosum agenesis,
KW   schizencephaly, diagnosis, assay, human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Sig_peptide
FT   Protein         24..1910
FT                   /label= Mat_protein
FT   Domain          24..887
FT                   /label= IG
FT                   /note= "immunoglobulin type-C2 domain"
FT   Domain          888..1594
FT                   /label= FbN
FT                   /note= "fibronectin type III domain"
FT   Domain          1595..1616
FT                   /label= Transmembrane
FT   Domain          1617..1910
FT                   /label= Cytoplasmic
FT   Region          24..126
FT                   /label= Ig1
FT   Region          127..225
FT                   /label= Ig2
FT   Region          226..316
FT                   /label= Ig3
FT   Region          317..409
FT                   /label= Ig4
FT   Region          410..506
FT                   /label= Ig5
FT   Region          507..603
FT                   /label= Ig6
FT   Region          604..697
FT                   /label= Ig7
FT   Region          698..792
FT                   /label= Ig8
FT   Region          793..887
FT                   /label= Ig9
FT   Disulfide-bond  46..102
FT   Disulfide-bond  145..197
FT   Disulfide-bond  246..293
FT   Disulfide-bond  335..385
FT   Disulfide-bond  428..484
FT   Disulfide-bond  525..575
FT   Disulfide-bond  617..669
FT   Disulfide-bond  711..766
FT   Disulfide-bond  809..865
FT   Disulfide-bond  1307..1359
FT   Modified-site   78..80
FT                   /note= "Asn is N-glycosylated"
FT   Modified site   106..108
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   470..472
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   487..489
FT                   /note= "Asn is N glycosylated"
FT   Modified-site   658..660
FT                   /note= "Asn is N glycosylated"
FT   Modified-site   666..668
FT                   /note= "Asn is N glycosylated"
FT   Modified-site   710..712
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   748..750
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   795..797
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   924..926
FT                   /note= "Asn is N glycosylated"
FT   Modified-site   1142..1144
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1160..1162
FT                   /note= "Asn is N glycosylated"
FT   Modified site   1250..1252
FT                   /note= "Asn is N glycosylated"
FT   Modified-site   1271..1273
FT                   /note= "Asn is N glycosylated"
FT   Modified-site   1324..1326
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   1341..1343
FT                   /note= "Asn is N glycosylated"
FT   Modified-site   1488..1490
FT                   /note= "Asn is N-glycosylated"
XX
PN   WO9817795-A1.
XX
PD   30-APR-1998.
XX
PF   23-OCT-1997;   97WO-US19547.
XX
PR   25-OCT-1996;   96US-0029322.
XX
PA   (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI   Korenberg JR;
XX
DR   WPI; 1998-271791/24.
DR   N-PSDB; AAV31981.
XX
```

```
PT   New isolated Down's Syndrome-cell adhesion molecule - used to
PT   develop products for detection, diagnosis and therapy of
PT   developmental and neurological abnormalities
XX
PS   Claim 2; Page 73-78; 109pp; English.
XX
CC   This polypeptide comprises Down syndrome-cell adhesion molecule
CC   DS-CAM1, a cell surface glycoprotein belonging to a novel subclass
CC   of the Ig superfamily with highest homology to neural cell adhesion
CC   molecules.  Its amino acid sequence was deduced from cDNA clones
CC   (see AAV31981) isolated from a trisomy 21 foetal brain library.  A
CC   splice variant, DS-CAM2 (see AAW42087), which is non-membrane bound
CC   was also identified.  The invention also provides human and murine
CC   DS-CAM nucleic acid sequences (see also AAV31985-88), expression
CC   vectors and host cells, transgenic animals, antibodies, antisense
CC   oligonucleotides, and primers derived from DS-CAM nucleic acid.
CC   DS-CAM polypeptides are associated with developmental and
CC   neurological processes.  They can be used in e.g. neural prosthetic
CC   devices used in entubulation methods of repairing (regenerating)
CC   damaged or severed peripheral nerves, and also in bioassays to
CC   identify agonists and antagonists.  The products can also be
CC   used in detection, diagnosis and therapy of developmental and
CC   neurological abnormalities such as Down syndrome, mental
CC   retardation, holoprosencephaly, agenesis of the corpus callosum,
CC   or schizencephaly.
XX
SQ   Sequence   1910 AA;


  Query Match             12.7%;  Score 170;  DB 19;  Length 1910;
  Best Local Similarity   29.8%;  Pred. No. 1.3e-05;
  Matches   57;  Conservative   31;  Mismatches   69;  Indels   34;  Gaps   10;

Qy      23 PREMTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYKEEGKVDLKSGNK- 81
           ||:: |||                 ||| || |:|   | ::|| |||  | : |  |
Db     320 prkvkssv---------------gsqvslscsvtg-tedqelswyr-ngei-lnpgknv 360

Qy      82 ----INSSSVCVSSISENDNGISFTCRLGRDQSVS---VSVVLNVTFPPLLSGNDFQTVE 134
               ||  .        | !    |   |       | |||    |  :|     :  |
Db     361 ritginhenlimdhmvksdgg-ayqcfvrkdklsaqdyvqvvledgtpkiisafsekvvs 419

Qy     135 EGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQ--TSES---FQLSITKVEKPD 189
             | |:|||  |   : |  :   : |:   |:| |  |||      |:|:  :  |
Db     420 paepvslmcnvkgtplptitwtldddpi-lkggshrisqmitsegnvvsylnisssqvrd 478

Qy     190 NGTYSCIAKSS 200
           | | | | :|
Db     479 ggvyrctanns 489


RESULT   5
AAW94996
ID   AAW94996 standard; Protein; 431 AA.
XX
AC   AAW94996;
XX
DT   12-MAY-1999  (first entry)
XX
DE   SAF-2 polypeptide (derived from EST).
XX
KW   Sialoadhesin; SAF-2; cellular interaction molecule; cancer; inflammation;
KW   autoimmunity; allergy; asthma; rheumatoid arthritis; multiple sclerosis;
KW   cerebellar degeneration; Alzheimer's disease; Parkinson's disease; AIDS;
KW   amylotrophic lateral sclerosis; head injury; neurological abnormality;
KW   septic shock; stroke; osteoporosis; osteoarthritis; ischemia reperfusion;
KW   cardiovascular disease; kidney; myocardial; hypotension; hypertension;
KW   myelodysplastic; anaemia; infection.
XX
OS   Homo sapiens.
XX
PN   CA2226910-A.
XX
PD   02-OCT-1998.
XX
PF   26-MAR-1998;   98CA 2226910.
XX
PR   02-APR-1997;   97US-0041886.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Erickson-Miller CL,  Kikly KK;
XX
DR   WPI; 1999-106564/10.
XX
PT   New sialoadhesin (SAF-2) polypeptide and polynucleotide - useful as
PT   diagnostic reagents and for prevention and treatment of cancer,
PT   AIDS, and Alzheimer's disease
XX
PS   Claim 14; Page 23-24; 38pp; English.
XX
CC   The invention relates to a sialoadhesin polypeptide (SAF-2) which is
CC   utilised as a cellular interaction molecule. SAF-2 polynucleotides and
CC   polypeptides are useful for diagnosing susceptibility to diseases and for
CC   screening for antagonists, agonists. They can also be used to treat
CC   diseases associated with a SAF-2 imbalance, including inflammation,
CC   autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,
CC   cerebellar degeneration, Alzheimer's disease, Parkinson's disease,
CC   multiple sclerosis, amylotrophic lateral sclerosis, head injury damage
CC   and other neurological abnormalities, septic shock, cancer, stroke,
CC   osteoporosis, osteoarthritis, ischemia reperfusion injury, kidney, liver,
CC   cardiovascular diseases, myocardial infarction, AIDS, hypotension,
CC   hypertension, myelodysplastic syndromes and other abnormalities, aplastic
CC   anaemia, male pattern baldness, and bacterial, fungal, protozoan and
CC   viral infections. SAF 2 polypeptides can be used as vaccines and in gene
CC   therapy to prevent the diseases. SAF 2 polypeptides are useful for
CC   mapping genes to chromosomes, allowing gene inheritance to be studied
CC   through linkage analysis.
XX
SQ   Sequence   431 AA;


  Query Match             12.7%;  Score 169.5;  DB 20;  Length 431;
  Best Local Similarity   23.6%;  Pred. No. 2e-06;
  Matches   63;  Conservative   53;  Mismatches   92;  Indels   59;  Gaps   11;

Qy      24 REMTSSVLTVNGKTENYILDT-TPGSQASLICAVQNHTREEELLWYREEGRVDLKS--GN 80
           ::::  |  :  : :  || |   |  :| |:|          |  ::|   : |  |
Db     144 kqlsvfvtalthrpdililgtlesghsrnltcsvp--------wackqgtppmiswiqa 194

Qy      81 KINS-------SSVCVSSISENDNGISFTCRL---GRDQSVSVSVVLNVTFPP------- 123
            ::|       |||   :    |:| | ||::   |   : : :| |:|:::|
Db     195 svsspgpttarssvltltpkpqdhgtsltcqvtlpgtgvtttstvrldvsyppwnltmtv 254

Qy     124 ---------LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQT 174
                    |     :| || :::||| | :|| |:: | : |   | |  ||     :
Db     255 fqgdatastalqngsslsvleqqslrlvcavnsnppparlswtrgs--ltlcpsr-----s 307

Qy     175 SESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPIIAACV--- 229
           |    | : :|   | | ::| |:::  :: :   | ::::  |   |:  :  | |
Db     308 snpgllelprvhvrdegeftcraqnaqgsqhislslslqneqtgtsrpvsqvtlaavqga 367

Qy     230 ----VIFLTLCFGLIARRKKIMKLCMK 252
               : ||: |   |     |:: | |
Db     368 gatalaflsfciifi-----ivrscrk 389


RESULT   6
AAW94995
ID   AAW94995 standard; Protein; 431 AA.
XX
AC   AAW94995;
XX
DT   12-MAY-1999  (first entry)
XX
DE   SAF-2 polypeptide.
XX
```

KW Sialoadhesin; SAF-2; cellular interaction molecule; cancer; inflammation;
KW autoimmunity; allergy; asthma; rheumatoid arthritis; multiple sclerosis;
KW cerebellar degeneration; Alzheimer's disease; Parkinson's disease; AIDS;
KW amylotrophic lateral sclerosis; head injury; neurological abnormality;
KW septic shock; stroke; osteoporosis, osteoarthritis, ischemia reperfusion;
KW cardiovascular disease; kidney; myocardial, hypotension, hypertension,
KW myelodysplastic; anaemia; infection.
XX
OS Homo sapiens.
XX
PN CA2226910-A.
XX
PD 02-OCT-1998.
XX
PF 26-MAR-1998; 98CA-2226910.
XX
PR 02-APR-1997; 97US-0041886.
XX
PA (SMIF) SMITHKLINE BEECHAM CORP.
XX
PI Erickson-Miller CL, Kikly KK;
XX
DR WPI; 1999-106564/10.
DR N-PSDB; AAX17772.
XX
PT New sialoadhesin (SAF-2) polypeptide and polynucleotide useful as
PT diagnostic reagents and for prevention and treatment of cancer,
PT AIDS, and Alzheimer's disease
XX
PS Claim 1; Page 23; 38pp; English.
XX
CC This represents a sialoadhesin polypeptide (SAF-2) which is utilised as
CC a cellular interaction molecule. SAF-2 polynucleotides and polypeptides
CC are useful for diagnosing susceptibility to diseases and for screening
CC for antagonists, agonists. They can also be used to treat diseases
CC associated with a SAF-2 imbalance, including cancer, inflammation,
CC autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,
CC cerebellar degeneration, Alzheimer's disease, Parkinson's disease,
CC multiple sclerosis, amylotrophic lateral sclerosis, head injury damage
CC and other neurological abnormalities, septic shock, sepsis, stroke,
CC osteoporosis, osteoarthritis, ischemia reperfusion injury, kidney, liver,
CC cardiovascular diseases, myocardial infarction, AIDS, hypotension,
CC hypertension, myelodysplastic syndromes and other abnormalities, aplastic
CC anaemia, male pattern baldness, and bacterial, fungal, protozoan and
CC viral infections. SAF 2 polypeptides can be used as vaccines and in gene
CC therapy to prevent the diseases. SAF 2 polypeptides are useful for
CC mapping genes to chromosomes, allowing gene inheritance to be studied
CC through linkage analysis.
XX
SQ Sequence 431 AA;

Query Match 12.7%; Score 169.5; DB 20; Length 431;
Best Local Similarity 23.6%; Pred. No. 2e-06;
Matches 63; Conservative 53; Mismatches 92; Indels 59; Gaps 11;

Qy 24 REMTSSVILTVNGKTENYILDT-TPGSQASLICAVQNHTREEELLWYREEGRVDLKS--GN 80
Db 144 kqlsvtvtallthrpdiiilqttlesqhsrnltcsvp wackqqtppmiswiqa 194

Qy 81 KINS SSVCVSSISENDNGISFTCKL---GRDQSVSVSVVLNVTFPP------- 123
Db 195 svssppqpttarss.i.ltpkpqpdhgtsltcqvtlpqtqvtttstvridvs;ppwnltmtv 254

Qy 124 ---------LLSGNDFQTVFFGSNVKIVCNVKANPQAQMMWYKNSSLLDLEKSPHQIQQT 174
Db 255 fqqdatastalqnqssisvlcqqslilvcavnstupparlswtrqs--ltlcpsr-----s 307

Qy 175 SPSPQESITKVEKPPNGTYSCIAKSSLKTESLDFRLIVKPKTVGV -PIEPTIAACV--- 229
Db 308 snpqlicipryhvrdeqgttcraqnaqqsqhislslqneqtqtsrpvsqvtlaavqqa 367

Qy 230 VIFLTLCFGLIAKKKKIMKLCMK 252

DB 368 gatalaflsfciiti -ivrserk 389

RESULT 7
AAY88565
ID AAY88565 standard; protein; 848 AA.
XX
AC AAY88565;
XX
DT 07-AUG-2000 (first entry)
XX
DE Human NCAM 140kD isoform precursor amino acid sequence.
XX
KW NCAM; neural cell adhesion molecule; Ig1; immunoglobulin domain 1;
KW neurite outgrowth promoter; proliferation; nerve damage; sclerosis;
KW impaired myelination, stroke, Parkinson's disease, memory, schizophrenia;
KW Alzheimer's disease, diabetes mellitus, circadian clock; nephrosis;
KW treatment, prosthetic nerve guide; treatment; nervous system.
XX
OS Homo sapiens.
XX
PN WO200018801-A2.
XX
PD 06-APR-2000.
XX
PF 23-SEP-1999; 99WO-DK00500.
XX
PR 29-SEP-1998; 98DK-0001232.
PR 29-APR-1999; 99DK-0000592.
XX
PA (RONN/) RONN L C B.
PA (BOCK/) BOCK E.
PA (HOLM/) HOLM A.
PA (OLSE/) OLSEN M.
PA (OSTE/) OSTERGAARD S.
PA (JENS/) JENSEN P H.
PA (POUL/) POULSEN F M.
PA (SORO/) SOROKA V.
PA (RALE/) RALETS I.
PA (BERE/) BEREZIN V.
XX
PI Ronn LCB, Bock E, Holm A, Olsen M, Ostergaard S, Jensen PH;
PI Poulsen FM, Soroka V, Ralets I, Berezin V;
XX
DR WPI; 2000-293111/25.
XX
PT Compositions that bind neural cell adhesion molecules useful for
PT treating disorders of the nervous system and muscles e.g. Alzheimer's
PT and Parkinson's diseases -
XX
PS Disclosure; Fig 17; 119pp; English.
XX
CC This sequence represents the human neural cell adhesion molecule (NCAM)
CC amino acid sequence. NCAM is found in three forms, two of which are
CC transmembrane forms, while the third is attached via a lipid anchor to
CC the cell membrane. All three NCAM forms have an extracellular structure
CC consisting five immunoglobulin domains (Ig domains). The Ig domains are
CC numbered 1 to 5 from the N-terminal. The invention relates to a compound
CC containing a peptide which binds to the NCAM Ig1 domain. The compound
CC binds to NCAM-Ig1/Ig2 domains, and is capable of stimulating or promoting
CC neurite outgrowth from NCAM presenting cells, and is also capable of
CC promoting the proliferation of NCAM presenting cells. The compound may be
CC used in the treatment of normal, degenerated or damaged NCAM presenting
CC cells. The compound may in particular be used to treat diseases of the
CC central and peripheral nervous systems such as post operative nerve
CC damage, traumatic nerve damage, impaired myelination of nerve fibres,
CC conditions resulting from a stroke, Parkinson's disease, Alzheimer's
CC disease, dementias, sclerosis, nerve degeneration associated with
CC diabetes mellitus, disorders affecting the circadian clock or
CC neuro muscular transmission and schizophrenia. Conditions affecting the
CC muscles may also be treated with the compound, such as conditions
CC associated with impaired function of neuromuscular connections

```
CC   (e.g. genetic or traumatic shock or traumatic atrophic muscle disorders).
CC   Conditions of the gonads, pancreas (e.g. diabetes mellitus types I and
CC   II), kidney (e.g. nephrosis), heart, liver and bowel may also be treated
CC   using the compound. The compound is used in a prosthetic nerve guide, and
CC   also to stimulate the ability to learn, and to stimulate the memory of a
CC   subject.
XX
SQ   Sequence   848 AA;


  Query Match             12.4%;  Score 166;  DB 21;  Length 848;
  Best Local Similarity   28.0%;  Pred. No. 1e-05;
  Matches   58;  Conservative   35;  Mismatches   80;  Indels   34;  Gaps    9;

Qy      49 QASLIC--------------AVQNHTREEELLWYREEGRVDLKSGNKINSSSVCVSSISE 94
           | :| |               : :| : ||: |    :| : ::|  ::  ||: : ||
Db     324 qvtltceasgdpipsitwrtstrnisseektl----dghmvvrsharv--ssltlksiqy 377

Qy      95 NDNGISFTC----RLGRDQSVSVSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQ 150
           | |  : |     :|:|   | |: | | : | | |       ||: | : | | | |
Db     378 tdag-eyictasntigqd---sqsmylevqyapklqgpvavytwegnqvnitcevfayps 433

Qy     151 AQMMWYKNSSLLDLEK-SRHQIQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFH 209
           | : |:::  ||     |  :|  |  :  | .|   . | | |.| | . .  |||.|
Db     434 atiswfrdgqllpssnysnikiyntpsasylevtpdsendfgnynctavnriqqeslefi 493

Qy     210 LIVKDKTVGVP----IEPIIAACVVIF 232
           |:  | |   |    :|| :   | |
Db     494 lvqad-tpsspsidqvepysstaqvqf 519


RESULT   8
AAB61142
ID   AAB61142 standard; Protein; 404 AA.
XX
AC   AAB61142;
XX
DT   30-MAR-2001  (first entry)
XX
DE   Human NOV12 protein.
XX
KW   Human, NOVX, antiinflammatory; cytostatic; neuroprotective;
KW   cerebroprotective; immunomodulator; vulnerary; vasotropic; gene therapy;
KW   hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
KW   diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
XX
OS   Homo sapiens.
XX
PN   WO200075321-A2.
XX
PD   14-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15305.
XX
PR   03-JUN-1999;   99US-0137322.
PR   16-MAR-2000; 2000US-0189810.
PR   22-MAR-2000; 2000US-0191158.
PR   30-MAR-2000; 2000US-0193086.
PR   31-MAY-2000; 2000US-0137322.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Fernandes E,  Herrman J,  Vernet C;
XX
DR   WPI; 2001-102403/11.
DR   N-PSDB; AAF27860
XX
PT   New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
PT   diagnostic marker, protein therapeutic, antibody or small molecule drug
PT   target for treating immune, proliferative and metabolic diseases and
PT   wound healing  -
XX
PS   Claim 1; Page 44-46; 194pp; English.
XX
CC   The present sequence is a new isolated polypeptide (NOVX). The NOVX
CC   polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
CC   treating or preventing NOVX associated disorders. They are also useful
CC   for determining the presence of or a predisposition to a disease
CC   associated with altered levels of the NOVX polypeptide or nucleic acid.
CC   These NOVX-associated disorders include hyperplasias, tumours,
CC   restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
CC   rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
CC   oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
CC   are especially useful in gene therapy. Specifically, NOVX is useful as
CC   a diagnostic marker or prognostic marker, protein therapeutic and
CC   antibody target or small molecule drug target to treat disorders in the
CC   immune response pathway, thyroid and metabolic diseases, bone metabolic
CC   disorders, diseases of the pancreas (e.g. diabetes or digestive
CC   disorders), proliferative diseases, or tissue regeneration and
CC   development (e.g. wound healing or treatment of burns).
XX
SQ   Sequence   404 AA;


  Query Match             12.1%;  Score 162;  DB 22;  Length 404;
  Best Local Similarity   22.4%;  Pred. No. 8.4e-06;
  Matches   64;  Conservative   48;  Mismatches  112;  Indels   62;  Gaps   13;

Qy      22 LPREMTSSVLTVNG-----KTENYILDTTPGSQASLICAVQNHTREEELLWYREEGRV-D 75
           :| : : : ||| |     :   :       |   | |       :: |:: : : |
Db     118 mpvktskaylfvlgvpgkpqisgfsspymegdlmqltcktsgskpaadirwfkndkeikd 177

Qy      76 LK--SGNKINSSSVCVSS-----ISENDNGISFTCRLGRDQ----------------SVS 112
           :|      |  :  |||      :  :|:|::  ||:  :                  : |
Db     178 vkylkeedanrktftvsstldfrvdrsddqvavicrvdheslnatpqvamqvleihytps 237

Qy     113 VSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQ-MMWYKN-SSLLDLEKSRHQ 170
           | :: :  ||           :||  : | |  |  |  : ::| |:   | | ::
Db     238 vkiipstpfp----------qegqpliltceskgkplpepvlwtkdqqelpdpdr---- 282

Qy     171 IQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTV----GVPIEPIIA 226
                 :|:|  : | ||||| | | :::   | :: ||| |         |   :|
Db     283 --mvvsgrelnilflnktdngtyrceatntiggssaeyvlivhdpnalagqngpdhalig 340

Qy     227 --ACVVIFLILC----FGLIARRKKIMKLCMKDKD----FHSETAL 262
             ||:|:|||     |    | |   |  : |     | ::||:
Db     341 givavvvfvtlcsifllgrylarhkgtyltneakgaedapdadtai 386


RESULT   9
ABB71337
ID   ABB71337 standard; Protein; 1079 AA.
XX
AC   ABB71337;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 40803.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   24-MAR-2000; 2000US-191537P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
```

```
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15440
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions
XX
PS   Disclosure; SEQ ID NO 40803; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1079 AA;


  Query Match             12.1%;  Score 161;  DB 22;  Length 1079;
  Best Local Similarity   24.3%;  Pred. No. 3.8e-05;
  Matches   46;  Conservative   34;  Mismatches   81;  Indels   28;  Gaps    6;

Qy      31 LTVNGKTENYILDTTPGSQASLICAVQNHTREEELLWYR-----EEGRVDL--------- 76
           : :||   :    |  ||:  | | :        | ||        |  ::
Db     265 idinqvvsh -- tvqqsevvltedikpirpavaltwynttttiissgeneiteyrsksle 320

Qy      77 KSGNKINSSSVCVSSISENDNGISFTC-------RLGRD-SVSVSVVLNVTFPPLLSGNE 129
           ||    :: |  : : :  :|   | |       :: ||: :| :: | | :||::  :
Db     321 ksdqttntqselifnatrtendrvtreeaeivlplarekpissaltlevlyppvvkesp 380

Qy     130 FQTVEEGSNVKLV CNVKANPQA  QMMWYKNSSLLDLEKSKHQ IQQTSESPQLSITKVE 186
           | : |: |   ||| :  |: ||:|  |:::  : |    ||:  | |  |
Db     381 saitantsgiclkneytarpsitqswvrtnlitvncnd'thpkggseevalvlkstq 440

Qy     187 KPDNGTYSC 195
           | | | |||
Db     441 kediqnysc 449


RESULT  10
AAY83741
ID   AAY83741 standard; Protein; 444 AA.
XX
AC   AAY83741;
XX
DT   09-NOV-1999  (first entry)
XX
DE   Beta-secretase.
XX
KW   Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;
KW   Alzheimer's disease.
XX
OS   Homo sapiens
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..24
FT                   /label= Signal_peptide
FT   Protein         24..444
FT                   /label= beta-secretase
FT   Region          377..399
FT                   /note= "Putative transmembrane region"
XX
PN   US5942400-A.
XX
PD   24-AUG-1999.
XX
PF   07-JUN-1996;   96US-0659984.
XX
PR   07-JUN-1996;   96US-0659984.
PR   07-JUN-1995;   95US-0480498.
PR   07-JUN-1995;   95US-0485152.
XX
PA   (ELAN-) ELAN PHARM INC.
XX
PI   Andersen JP, Jacobsen Creak KL, Sinha S;
XX
DR   WPI; 1999-517417/43.
DR   N-PSDB; AAZ06640.
XX
PT   A method for detecting human beta-secretase cleavage of polypeptides
PT   useful for identifying beta-secretase inhibitors
XX
PS   Examples; Fig 1; 43pp; English.
XX
CC   This sequence is the human beta-secretase enzyme polynucleotide.
CC   Beta-secretase is capable of cleaving the beta-amyloid precursor protein
CC   (APP) (AAY83742,swedish mutant APP). This enzyme is used in a method for
CC   detecting human beta secretase cleavage of polypeptides and for
CC   identifying beta-secretase inhibitors. The protein has a deduced
CC   molecular weight of 49 kD, although this is without glycosylation and
CC   the expected molecular weight is therefore higher. Inhibition of
CC   beta-secretase activity would be useful for chemical modelling of a
CC   critical event in the pathology of Alzheimer's disease. Inhibitors of
CC   beta-secretase would be useful for the prevention and treatment of
CC   Alzheimer's disease and Down's Syndrome.
XX
SQ   Sequence   444 AA;


  Query Match             11.8%;  Score 158;  DB 20;  Length 444;
  Best Local Similarity   20.6%;  Pred. No. 2.1e-05;
  Matches   62;  Conservative   48;  Mismatches  109;  Indels  102;  Gaps   13;

Qy      22 LPFEMISFVLIVNGKIE-----NYILDTTPGSQASLICAVQNHTPEEFILWYPFFGRV-D 75
           :| : : : ||| |  |     :     |   | |          :: |:: :  : |
Db     118 apvktskayltvlgvpkkpgisqfssqvmegdlmqpdktsqskpaadirwyladkeikd 177

Qy      76 LK--SGNKINSSSVCVSS-----ISENDNGISFTCRLGRDQ----------------SVS 112
           :|       |  :  |||      :  :|:|::  ||:  :                 : |
Db     178 vkylkeedanrktftvsstldfrvdrsddqvavicrvdhesinatpqvamqvleihytps 237

Qy     113 VSVVLNVTFPPLLSGNDFQIVEEGSNVKLVCNVKANPQAQ-MMWYKN-SSLLDLEKSKHQ 170
           | :: :  ||           :||  : | |  |  |  : ::| |:   | | ::
Db     238 vkiipstpfp           qeqqpliltcesvgkplpepvlwtkdgqelpdpdr---- 282

Qy     171 IQQTSESPQLSITKVEKEDNGTYSCIAKGSLETESLDFKLIVKD                  214
               :|:|  : | ||||| | | :::   | :: ||| |
Db     283   mvvsyrelnilflnktdngtyrceatntiqqssaeyvlivhdvpntllptiipsltt 340

Qy     215 ----KTVCVPIEPIIAA-------------------------CVVIFLTLC----FGLI 240
                 || :   |  :|                         |||:|||     |
Db     341 atvtttvaittsptsatssirdpnalaqqnqpdhaliqqivavvvfvtlesiflqry 400

Qy     241 ARRKKIMKLCMKDKD----PHSETAL 262
           | |   |  : |     | ::||:
Db     401 larhkqtyltneakqaedapdadtai 426


RESULT  11
AAB47251
ID   AAB47251 standard; Protein; 444 AA.
XX
AC   AAB47251;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Beta-secretase.
XX
KW   Beta-secretase, isotype, beta amyloid precursor protein; APP;
```

```
KW   beta-amyloid peptide; beta-AP; Alzheimer's disease; Downs syndrome;
KW   HCHWA-D; Swedish mutation.
XX
OS   Homo sapiens
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /note= "Putative signal peptide"
FT   Protein         24..444
FT                   /note= "Mature beta-secretase"
FT   Peptide         125..141
FT                   /label= Peptide 1
FT   Peptide         190..200
FT                   /label= Peptide 5
FT   Peptide         240..259
FT                   /label= Peptide 2
FT   Peptide         261..272
FT                   /label= Peptide 3
FT   Domain          377..399
FT                   /note= "Putative transmembrane domain"
XX
PN   US6221645-B1.
XX
PD   24-APR-2001.
XX
PF   07-JUN-1996;   96US-0660531.
XX
PR   07-JUN-1995;   95US-0480498.
XX
PA   (ELAN-) ELAN PHARM INC.
XX
PI   Chrysler SMS,  Sinha S,  Keim PS,  Anderson JP,  Tan H,  McConlogue L;
XX
DR   WPI; 2001-315578/33.
DR   N-PSDB; AAC85809.
XX
PT   Novel antibody that specifically binds native beta-secretase protein,
PT   useful for raising anti-idiotypic antibodies and for detecting or
PT   diagnosing pathological conditions related to presence of respective
PT   antigens -
XX
PS   Example; Fig 1; 42pp; English.
XX
CC   This sequence represents beta-secretase protein which is characterized
CC   by an ability to cleave the 695-amino acid isotype of beta-amyloid
CC   precursor protein (APP) between amino acids 596 and 597.  Beta-
CC   secretase is thought to be responsible for the pathogenic processing
CC   of APP to form beta-amyloid peptide (beta-AP) in beta-AP related
CC   conditions, e.g. Alzheimer's disease, Downs syndrome, HCHWA-D etc.
CC   Beta-secretase has a molecular weight of 260-300 kD and will bind
CC   to wheat germ agglutinin but not to concanavalin A.  Beta-secretase
CC   will cleave both the wild type and the Swedish mutation of APP.
XX
SQ   Sequence   444 AA;


  Query Match           11.8%;  Score 158;  DB 22;  Length 444;
  Best Local Similarity  20.6%;  Pred. No. 2.1e-05;
  Matches   67;  Conservative   48;  Mismatches  109;  Indels  102,  Gaps   13,

Qy      22 LPREMTSSVLTVNGKTE-----NYILDTTPGSQASLICAVQNHTREEELLWYREEGRV-D 75
           :| : : : ||| |  |       :     |    | |          :: |:: :  : | 
Db     118 mpvktskayltvlgvpekpqisgfsspvmegdlmqltcktsgskpaadirwfkndkeikd 177

Qy      76 LK--SGNKINSSSVCVSS-----ISENDNGISFTCRLGRDQ---------------SVS 112
           :|       |  :  |||     :  :|:|::  ||:  :                 : | 
Db     178 vkylkeedanrktftvsstldfrvdrsddgvavicrvdheslnatpqvamqvleihytps 237

Qy     113 VSVVLNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQ-MMWYKN-SSLLDLEKSRHQ 170
           | :: :  ||          :||  : | |  |  |  : ::| |:   | | ::
Db     238 vkiipstpfp----------qegqpliltceskgkplpepvlwtkdggelpdpdr---- 282

Qy     171 IQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKD--------------- 214

                :|:|  : | ||||| | | :::   | :: ||| |
Db     283 --mvvsgrelnilfinktdngtyrceatntiqqssaeyvlivhdvpntiiptiipsltt 340

Qy     215 ----KTVGVPIEPIIAA-------------------------CVVIFLTLC----FGLI 240
                 || :   |  :|                          ||:|:|||     |
Db     341 atvtttvaittspttsattssirdpnalagqngpdhaliggivavvvfvtlcsifllqry 400

Qy     241 ARRKKIMKLCMKDKD----PHSETAL 262
           | |   |  : |     | ::||:
Db     401 larhkgtyltneakgaedapdadtai 426


RESULT  12
AAB50692
ID   AAB50692 standard; Protein; 1415 AA.
XX
AC   AAB50692;
XX
DT   19-MAR-2001  (first entry)
XX
DE   C. elegans UNC-40 protein SEQ ID NO:93.
XX
KW   Human; Caenorhabditis elegans; UNC-5; splice variant; nematode worm;
KW   protein-protein interaction; identification.
XX
OS   Caenorhabditis elegans.
XX
PN   WO200073328-A2.
XX
PD   07-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-EP05108.
XX
PR   01-JUN-1999;   99GB-0012755.
XX
PA   (DEVG-) DEVGEN NV.
XX
PI   Van Criekinge W,  Roelens I,  Bogaert T,  Verwaerde P;
XX
DR   WPI; 2001-016508/02.
DR   N-PSDB; AAC90960.
XX
PT   Three variants of human unc-5C cDNAs (unc-5Cb, unc-5Cc and unc-5C8) and
PT   a human unc-5HS1 cDNA, useful in yeast two hybrid experiments for
PT   identifying unknown human cDNAs which encode proteins that interact
PT   with the human unc-5C protein -
XX
PS   Disclosure; Page 233-238; 246pp; English.
XX
CC   The present invention describes 3 variants of human unc-5C cDNAs
CC   (unc-5Cb, unc-5Cc and unc-5C8) which correspond to alternatively spliced
CC   unc-5C transcripts, and a human unc-5HS1 cDNA which shares homology with
CC   the Rattus norvegicus unc-5HS1 cDNA. Also described are assays based on
CC   protein-protein-interactions between the unc-5 protein and a variety of
CC   different interacting proteins. The unc-5C variant cDNAs and unc-5HS1
CC   cDNA are useful in methods for identifying compounds which reduce or
CC   inhibit the lethal phenotype associated with the expression of the
CC   unc-5 death domain in yeast. They are also useful in yeast two hybrid
CC   experiments for identifying unknown human cDNAs which encode proteins
CC   that interact with the human unc-5C protein. AAC90914 to AAC90971 and
CC   AAB50646 to AAB50693 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   1415 AA;


  Query Match           11.8%;  Score 157.5;  DB 22;  Length 1415;
  Best Local Similarity  27.8%;  Pred. No. 0.00011;
  Matches   58;  Conservative   35;  Mismatches   83;  Indels   33;  Gaps   10;

Qy      24 REMTSSVLTVNGKT--ENYILDTTP-------GSQASLICAVQNHTREEELLWYREEGR- 73
           |   :: |||  :|     :  |||       | :  | | | :  |  :: | ::  :
Db     227 rssqtarltvttetvsnelvfittprlqvveqgdefileclvaslir pqrrwlkdsrqi 285
```

```
Qy      74 -VDLKSGNKINSSSVCVSSISPNDNGISEDCRI--GRDQSVSVSVVLNVIEPPELSGNDFQ 131
           ||      ::  ||: ||  |  | |: :|||    |||:  :| : |  || ::
Db     286 ivdpvsirvqgssilvsrasseditl ,lrrasnoddsidrassevraqpriffqptt 344

Qy     132 TVFCGSNVKLVC--NVKANPQAQMMWYKNSSLLDEKSPDQIQQISESFSQLSTLTVEKPDNGIYSCIAKSS 200
           |  | ::|:| |    | |:|:: |||     |   :| :|           | |
Db     345 kvazetadveleeqtaairpearvnwykngeiiqseyfviepar ---       lril 393

Qy     184 KVEEPDNGTYPCIAKSCLETESLDFHLIV 212
           | : |   | |||:: : :|     |:|
Db     394 qvvradqaiyqeiaendvqsesasaqllv 422


RESULT  13
ABB63044
ID   ABB63044 standard; Protein; 467 AA.
XX
AC   ABB63044;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 15924.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07147.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 15924; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL33511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57747-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   467 AA;


  Query Match             11.8%;  Score 157;  DB 22;  Length 467;
  Best Local Similarity   26.5%;  Pred. No. 2.8e-05;
  Matches   48;  Conservative   37;  Mismatches   72;  Indels   24;  Gaps    8;

Qy      42 TVNGKTENYILDTTPGSQASLICAVQNHTREFEFLWYREEGRVDLKSG--NKINSSSVCV 89
           |:     |  :    || :| |    :     : |:::    |: ||  :  :|||:: :
Db     146 tiralphnqqvtarkqstvtlockasgnp vptifwfkk    dvfsgpthlsdsstlil 190
```

```
Qy      90 SSISFNDNGISEDCRI--GRDQSVSVSVVLNVIEPPELSGNDFQIVE-------EGSNVK 140
           ::  :  | :: |    |    ||: : | :  || :          |||       || :|:
Db     191 ravdrkbagitypsadeq-kdrvsmdigiilspeei--   -tvekswchasedydre 243

Qy     141 LVCNVKANPQAQMMWYKNSSLLDEKSPDQIQQISESFSQLSTLTVEKPDNGIYSCIAKSS 200
           ||| |  :  ::|:|||:|| |||    |  :    : : | |   :  | | |||:| ::
Db     244 lvsivhqdvnsemlwygnsfild atdrrsmyprddrysliirntqptdtqnyscvadna 302

Qy     201 L 201
           |
Db     303 I 303


RESULT  14
AAB70070
ID   AAB70070 standard; Protein; 595 AA.
XX
AC   AAB70070;
XX
DT   14-MAY-2001  (first entry)
XX
DE   Human secreted protein #9.
XX
KW   Human; secreted protein; immunomodulatory; antisclerotic;
KW   dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW   immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW   ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW   antialzheimers; antiparkinsonian; antimicrobial; vulnerary; gene therapy;
KW   immune disorder; hyperproliferative; cardiovascular; angiogenic;
KW   neurological; infection.
XX
OS   Homo sapiens.
XX
PN   WO200112776-A2.
XX
PD   22-FEB-2001.
XX
PF   15-AUG-2000; 2000WO-US22350.
XX
PR   16-AUG-1999;   99US-0148759.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Shi Y,  Young PE,  Ebner R,  Soppet DR,  Ruben SM;
XX
DR   WPI; 2001-244245/25.
DR   N-PSDB; AAF76851.
XX
PT   Nucleic acids encoding 18 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Claim 11; Page 364-366; 380pp; English.
XX
CC   The present sequence is one of 18 novel human secreted proteins. The
CC   nucleic acids encoding the proteins and the proteins themselves may be
CC   used in the prevention, diagnosis and treatment of diseases including
CC   immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
CC   and human immunodeficiency virus (HIV) infections), hyperproliferative
CC   disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC   (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC   arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC   neovascularisation and diabetic retinopathy), neurological disorders
CC   (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC   infectious diseases and/or for promoting wound healing, regeneration
CC   and/or chemotaxis. The nucleic acid molecules may also be used as DNA
CC   probes in diagnostic assays to detect and quantitate the presence of
CC   similar nucleic acid sequences in samples. The polypeptides may also be
CC   used as antigens in the production of antibodies and in assays to
CC   identify modulators of protein expression and activity.
XX
SQ   Sequence   595 AA;
```

```
  Query Match             11.6%;  Score 155;  DB 22;  Length 595;
  Best Local Similarity   23.2%;  Pred. No. 5.7e-05;
  Matches   63;  Conservative   46;  Mismatches   84;  Indels   78;  Gaps   12;

Qy      39 NYILD--------------TTPGSQAS-----LICAVQNHTREEELLWYREEGR----- 73
           ||| |              : ||:  |     | |:|          |  |:|
Db     258 nyiydklsvhvtalthmptfsipgtlesghprnltcsvp---------waceqgtpptit 308

Qy      74 ---VDLKSGNKINSSSVCVSSISE-NDNGISFTCRL---GRDQSVSVSVVLNVTFPP--- 123
             : | :   : |  :| | :  |:| | ||::   |   ::: :| ||:::||
Db     309 wmgasvssidptitrssmlslipqpqdhgtsltcqvtlpgagvtmtravrlnisyppqnl 368

Qy     124 ------------LLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQ 170
                       |      :| || :: ||| | :|| |:: |   |  | |  |
Db     369 tmtvfqgdgtasttlrngsalsvleggslhlvcavdsnppar1swtwgs--ltlsps--- 423

Qy     171 IQQTSESFQLSITKVEKPDNGTYSCIAKSSLKTESLDFHLIVKDKTVGV--PIEPI---- 224
             |:|    | : :|   | | ::| |:: | :: :   | :::: |   ||  :
Db     424 --qssnlgvlelprvhvkdegcftcraqnplgsqhislslslqneytgkmrpisgvtlga 481

Qy     225 ---IAACVVIFLTLCFGLIARRKKIMKLCMK 252
               |  ::||  |   :     ::: | |
Db     482 fggagatalvflyfciifv-----vvrscrk 507


RESULT  15
AAG77857
ID   AAG77857 standard; Protein; 537 AA.
XX
AC   AAG77857;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Mutant protein mMuSK-RI.
XX
KW   Mutated muscle specific tyrosine kinase receptor; selectable marker;
KW   hMuSK-R; human; mMuSK-R; antibody; mutant; mutein; immunoselection.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "Signal peptide of hMuSK-R"
FT   Domain          20..493
FT                   /note= "Extracellular domain"
FT   Domain          494..515
FT                   /note= "Transmembrane domain"
XX
PN   WO200172834-A1.
XX
PD   04-OCT-2001.
XX
PF   28-MAR-2001; 2001WO-EP03543.
XX
PR   30-MAR-2000; 2000US-0539248.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Pippig SD,  Veres G;
XX
DR   WPI; 2001-626257/72.
XX
PT   Identifying genetically modified mammalian cells comprising a nucleic
PT   acid sequence encoding a mutated muscle-specific tyrosine kinase
PT   receptor provides a selectable marker for transformation  -
XX
PS   Claim 5; Page-; 52pp; English.
XX
CC   The sequence represents the mutant protein mMuSK-RI. The invention
CC   concerns a method of using a muscle specific tyrosine kinase receptor
CC   molecule (MuSK-R) or a mutated MuSK-R (mMuSK-R) thereof as a selectable
CC   marker in mammalian cells, particularly human cells. Preferred markers
CC   are mMuSK-R's incapable of signal transduction and preferably molecules
CC   wherein the intracellular domain has been modified by deletion of the
CC   signalling region. The invention also relates to a method of identifying
CC   genetically modified mammalian cells including introducing a mMuSK-R into
CC   a target cell as a selectable marker. Further, a method for the
CC   immunoselection of transduced mammalian cells is disclosed comprising
CC   identifying the transduced cells by incubation of the cells with an
CC   antibody which recognises and binds specifically to a MuSK-R or mMuSK-R
CC   thereof.
CC   Note: The present sequence is not shown in the specification but is
CC   derived from the human wild-type muscle specific tyrosine kinase receptor
CC   protein shown in Figure 2 (AAG77856).
XX
SQ   Sequence   537 AA;


  Query Match             11.6%;  Score 154.5;  DB 22;  Length 537;
  Best Local Similarity   25.2%;  Pred. No. 5.4e-05;
  Matches   56;  Conservative   41;  Mismatches   82;  Indels   43;  Gaps   11;

Qy      15 LLLVILF-----LPRE--MTSSVLTVNGKTENYILDTTPGSQASLICAVQNHTREEELLW 67
           :| :: |     ||:   :|: : ||:   |           |: :|||:::  : |: |
Db      12 iltlvafsgteklpkapvittpletvdalveev---------atfmcavesyp-qpeisw 61

Qy      68 YREEGRV---DLKSGNKINSSSVCVSSISENDNGISFTCRLGRD--------QSVSVSVV 116
             | :  :   | :   : |   : : |: ::|:|| : |              :: | :
Db      62 trnkiliklfdtrysirengqlltilsvedsddgi-ycctanngvggavescgalqvkmk 120

Qy     117 LNVTFPPLLSGNDFQTVEEGSNVKLVCNVKANPQAQMMWYKNSSLLDLEKSRHQIQQTSE 176
           :| ||:       : ||    | |    ||:  : | |  | |  | ||  :    |
Db     121 pkitrppi-----nvkiieglkavlpcttmgnpkpsvswikgdspl-rensriav---le 171

Qy     177 SFQLSITKVEKPDNGTYSCIAKSSL-----KTESLDFHLIVK 213
           |  | |  |:| | | | |:||:||     |   |:| :  :
Db     172 sgslrihnvqkedagqyrcvaknslgtayskvvklefevfar 213


Search completed: September 17, 2002, 13:09:56
Job time: 11323 sec
```